How threatening are transformations of happiness scales to subjective wellbeing research?

Maastricht University

Caspar Kaiser, Maarten C.M. Vendrik

How threatening are transformations of happiness scales to subjective wellbeing research?

RM/20/032

ISSN: 2666-8807

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

How threatening are transformations of happiness scales to subjective wellbeing research? †

Caspar Kaiser* and Maarten C.M. Vendrik**

September 2020

Abstract

Two recent papers argue that many results based on ordinal reports of happiness can be reversed with suitable monotonic increasing transformations of the associated happiness scale (Bond and Lang 2019; Schröder and Yitzhaki 2017). If true, empirical research utilizing such reports is in trouble. Against this background, we make four main contributions. First, we show that reversals are fundamentally made possible by explanatory variables having heterogenous effects across the distribution of happiness. We derive a simple test of whether reversals are possible by relabelling the scores of reported happiness and deduce bounds for ratios of coefficients under any labelling scheme. Second, we argue that in cases where reversals by relabelling happiness scores are impossible, reversals using an alternative method of Bond and Lang, which is based on ordered probit regressions, are highly speculative. Third, we make apparent that in order to achieve reversals, the analyst must assume that respondents use the response scale in a strongly non-linear fashion. However, drawing from the economic and psychological literature, we present arguments and evidence which suggest that respondents likely use response scales in an approximately linear manner. Fourth, using German SOEP data, we provide additional empirical evidence on whether reversals of effects of standard demographic variables are both possible and plausible. It turns out that reversals by either relabelling or by using Bond & Lang's approach are impossible or implausible for almost all variables of interest. Although our analysis uses happiness as a special case, our theoretical considerations are applicable to any type of subjective ordinal report.

JEL Codes: I31, C25

Keywords: ordinal reports, transformations of cardinal scales, happiness, subjective wellbeing, life satisfaction, Easterlin Paradox, General Social Survey, German Socio-Economic Panel

---

† This paper is based on an earlier manuscript available on the Open Science Framework with a slightly different title (see previous versions on https://osf.io/gzt7a/). We thank Timothy Bond and Kevin Lang for kindly sharing their replication files. We further thank Jan-Emmanuel De Neve, Martijn Hendriks, Brian Nolan, Andrew Oswald, Michael Plant, Alberto Prati, as well as participants of the Oxford Wellbeing Research Centre seminar series and the ISQOLS 2019 (Granada) Conference for helpful comments and suggestions. Funding from Nuffield College, the Department of Social Policy & Intervention, and the Wellbeing Research Centre (Oxford) is gratefully acknowledged.

* Nuffield College, Department of Social Policy & Intervention, Wellbeing Research Centre, University of Oxford. Email: caspar.kaiser@nuffield.ox.ac.uk

** Department of Macro, International and Labour Economics and ROA, SBE, Maastricht University; IZA, Bonn; EHERO, Erasmus University, Rotterdam. Email: m.vendrik@maastrichtuniversity.nl

1 Introduction

This paper offers an analysis of the conditions under which reports about happiness can be used to identify the effects of socio-economic variables on actual "true" happiness. Some of this analysis is a reaction to a recent paper by Bond and Lang (2019). In that paper, Bond and Lang forcefully argue that the results of most happiness research can be reversed, i.e. that the estimated sign of a variable's effect on mean happiness can be inverted by means of some suitable transformation of the cardinal happiness scale under consideration. In a similar vein, Schröder and Yitzhaki (2017) show that signs of coefficients from OLS regressions of reported happiness are potentially reversible. If Bond and Lang and Schröder and Yitzhaki are right, happiness research is in trouble. It is therefore imperative to give a comprehensive analysis of the reasons why and the conditions under which such reversals are possible.

In giving such an analysis we make four key points. First, we show that sign reversals are fundamentally made possible by explanatory variables having heterogenous effects across the distribution of happiness. In the context of OLS regressions, we derive a simple test of whether reversals are possible by relabelling the scores of reported happiness. Second, we argue that in cases where reversals by relabelling happiness scores are impossible, reversals based on Bond and Lang's method using ordered probit regressions are empirically unfounded and thus highly speculative. Third, we make apparent that in order to achieve reversals, the analyst must assume that respondents use the response scale in a strongly non-linear fashion. However, drawing from the economic and psychological literature, we provide arguments and discuss experimental evidence which suggest that respondents likely use response scales in an approximately linear manner. Deviations from linearity seem to be larger for response scales with as few as three response options. Fourth, using German SOEP data, we provide additional empirical evidence on whether reversals of effects of standard demographic variables are both possible and plausible. It turns out that reversals by relabelling are impossible or implausible for almost all variables of interest. Likewise, when using Bond and Lang's approach, reversals always rely on assuming implausible ways in which respondents use the response scale.

Before making these and some additional points, it is helpful to define our terms and to frame the discussion. Happiness is typically measured via responses to questions like *"How happy are you these days?"* or *"Taking all things together, how satisfied are you with your life?"*.[1] Such responses are then recorded in their rank order, i.e. giving a "1" to the first response option, a "2" to the second response option, and so on. Call these ordered responses hr (**h**appiness **r**eports) with options $\{1,2,\dots,R\}$. These happiness reports are assumed to be indicative of an unobservable cardinal quantity of true happiness, which is a subjective feeling whose intensity is only internally accessible to the respondent. Call this quantity ht (**h**appiness **t**rue). We take it that most research on happiness is concerned with estimating the direction and magnitude of effects of socio-economic variables on expected quantities of true happiness ht as approximated by ordered categories of reported happiness hr. Note that we exclusively focus on issues that arise even if one assumes that scale use is homogenous across all respondents and times. We thus ignore the additional problems arising from inter- and intra-personal differences in scale use (see e.g. King et al. 2004).

[1] Responses to the latter question are often taken to measure life satisfaction rather than happiness. Since this distinction makes no difference to the arguments of the present paper, we will primarily use the term "happiness" in a wider sense than merely affective wellbeing.

The paper is structured as follows. In Section 2 we analyse why reversals are possible, derive conditions under which reversals by relabelling become impossible, and provide bounds on trade-off ratios of coefficients under any labelling scheme. In Section 3 we explore the implications of allowing ht to vary with explanatory variables within each response category of hr, analyse Bond and Lang's approach of achieving reversals in the context of ordered probit regressions, and give a comparison of Bond and Lang's and our approach. In Section 4 we note that reversals typically require that respondents use the response scale in a strongly non-linear fashion. We then present several arguments and experimental evidence showing that respondents likely use the response scale in a roughly linear fashion. Section 5 provides further empirical evidence using German SOEP data. A final section concludes.

2 Why are sign reversals possible?

In this section we analyze under which conditions sign reversals of effects of changes in explanatory variables on happiness are possible and derive bounds on trade-off ratios of coefficients under any labelling scheme.

2.1 Easterlin Paradox example

To introduce the issue, consider the following example. The example concerns the validity of the Easterlin Paradox, which states that there is no long-term effect of changes in *per capita* income on mean happiness over time. This particular example is also given in Section A.3.1 of Bond and Lang (2019). It thereby allows for a direct comparison between their and our analysis. The example is ultimately based on an analysis by Stevenson and Wolfers (2008) in which happiness data from the US-American General Social Survey are regressed on national *per capita* income data from the Federal Reserve Bank of St. Louis (spanning the period 1973-2006). The happiness data are individual answers to the survey question "*Taken all together, how would you say things are these days – would you say that you are very happy, pretty happy, or not too happy?*". To estimate the relation between mean happiness and real GDP *per capita* in the USA, Stevenson and Wolfers as well as Bond and Lang use an ordered probit regression.

However, a more straightforward and widely used approach is to code the three response categories of the happiness question as 1 for "not too happy", 2 for "pretty happy", and 3 for "very happy" and to run an OLS regression of this variable or its mean. The implicit assumption then is that the resulting rank-order scale (1, 2, 3) of reported happiness hr is a good approximation of the average cardinal values of the underlying "true" felt happiness intensity ht (on a continuous scale) within the three response categories. More precisely, it is presumed that within each category, the sample mean $E(ht|hr = k)$ equals k for $k = 1, 2, 3$.[2]

Mean happiness in the USA in a particular year is then estimated as mean happiness $E(ht)$ in the GSS sample, which is easily calculated as $s_1 * 1 + s_2 * 2 + s_3 * 3$, where s_k = the share of the sample that reports $hr = k$. Hence, when mean happiness $E(ht)$ is linearly related to an explanatory variable X (*e.g.*, the log of GDP per capita), the effect of a unit change in X on $E(ht)$

[2] More generally, a positive linear transformation $E(ht|hr = k) = a + b * k$ with $b > 0$ of this scale would yield identical signs and ratios of coefficients in a regression of $E(ht|hr = k)$. For the sake of convenience, we use the expectation notation E for sample as well as population means throughout this paper.

is given by $\partial E(ht)/\partial X = 1 * \partial s_1/\partial X + 2 * \partial s_2/\partial X + 3 * \partial s_3/\partial X$. Taking $hr = 2$ as the reference case, this can also be written as

$$\frac{\partial E(ht)}{\partial X} = (1-2) * \frac{\partial s_1}{\partial X} + (3-2) * \frac{\partial s_3}{\partial X} + 2 * \frac{\partial (s_1 + s_2 + s_3)}{\partial X} = \frac{\partial s_3}{\partial X} - \frac{\partial s_1}{\partial X}. \qquad (1)$$

In the case of the dataset of Stevenson and Wolfers and their test of the Easterlin Paradox for the USA (as adopted by Bond and Lang), $X = \ln GDPpc$. In that data, the share of the "very happy" (s_3) and that of the "not too happy" (s_1) fell with increasing $\ln GDPpc$ (see the OLS estimates in Table 1). The last two derivatives in Equation (1) are therefore both negative. Hence, the sign of $\partial E(ht)/\partial \ln GDPpc$ depends on the relative sizes of $\partial s_3/\partial \ln GDPpc$ and $\partial s_1/\partial \ln GDPpc$. Table 1 shows that the sign of $\partial E(ht)/\partial \ln GDPpc$ is negative (confirming the Easterlin Paradox for the USA) because $\partial s_3/\partial \ln GDPpc$ is more strongly negative than $\partial s_1/\partial \ln GDPpc$. In other words, the share of the "very happy" fell more strongly with economic growth than the share of the "not too happy".

Table 1. Regressions of shares and means of reported happiness on *per capita* GDP

	s_1	s_2	s_3	$\mathrm{E}(ht)$	$\widetilde{\mathrm{E}(ht)}$	$\widetilde{\widetilde{\mathrm{E}(ht)}}$	$\widetilde{\widetilde{\widetilde{\mathrm{E}(ht)}}}$
h_1, h_2, h_3				1, 2, 3	1, 2, 2.47	1, 2, 2	1, 2, 5.43
$(h_2 - h_1)/(h_3 - h_2)$					2.11	∞	0.29
ln*GDPpc*	-0.025	0.079***	-0.054**	-0.028	0.000	0.025	-0.158**
	(0.016)	(0.020)	(0.020)	(0.030)	(0.022)	(0.016)	(0.076)

Note: * p < 0.10; ** p < 0.05, *** p < 0.01. Rows for ln*GDPpc* denote regression coefficients with ordinary standard errors in parentheses. We obtain no significant serial correlation in the error of any regression and no significant heteroscedasticity for s_1, s_2, and $\widetilde{\widetilde{\mathrm{E}(ht)}}$. We do find significant heteroscedasticity for s_3, $\mathrm{E}(ht)$, $\widetilde{\mathrm{E}(ht)}$, and $\widetilde{\widetilde{\widetilde{\mathrm{E}(ht)}}}$, but in these cases heteroscedasticity-robust standard errors are smaller than, or very similar to ordinary standard errors (see Angrist and Pischke, 2009, p. 307 for details).[3]

However, the rank-order scale (1, 2, 3) for reported happiness hr, although intuitively plausible, may not correctly reflect the average difference in the underlying "true" felt happiness intensity ht between the three response categories. If so, $E(ht|hr = k)$ will not be equal to k for $k = 1, 2, 3$. For example, the difference between "pretty happy" and "not too happy" may be considerably larger in terms of average true happiness intensity $E(ht|hr = k)$ than the difference between "very happy" and "pretty happy". Denoting $E(ht|hr = k)$ as h_k for $k = 1, 2, 3$, the difference $h_2 - h_1$ would then be considerably larger than $h_3 - h_2$. This implies an alternative coding scale $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3)$ for reported happiness hr in the three response categories which is concave in the rank-order scale $(1, 2, 3)$ of hr.

For this alternative scale, Equation (1) becomes

$$\frac{\partial \widetilde{E(ht)}}{\partial X} = \left(\tilde{h}_3 - \tilde{h}_2\right)\frac{\partial s_3}{\partial X} - \left(\tilde{h}_2 - \tilde{h}_1\right)\frac{\partial s_1}{\partial X}, \qquad (2)$$

where $\widetilde{E(ht)} = \sum_{k=1}^{3} s_k * \tilde{h}_k$. Because $\tilde{h}_2 - \tilde{h}_1 > \tilde{h}_3 - \tilde{h}_2$, the fall of the share of the "not too happy" with increasing $\ln GDPpc$ ($\partial s_1/\partial \ln GDPpc$) will get a higher weight in the change of mean

[3] We also estimated all regressions with Feasible GLS and ML with simultaneous estimation of the standard deviation of the error. This yielded coefficient estimates for ln*GDPpc* that were suspiciously different in size from the OLS estimates, which suggests that the errors and ln*GDPpc* are correlated, implying inconsistency of the estimators (Wooldridge 2009, p.286). We therefore stick to the OLS estimates in Table 1.

happiness $\widetilde{E(ht)}$ relative to the fall of the share of the "very happy" ($\partial s_3/\partial \ln GDPpc$). With a sufficiently higher weight, this may reverse the sign of $\partial\widetilde{E(ht)}/\partial \ln GDPpc$ from negative to positive.[4] The ratio $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2)$ beyond which such a sign reversal occurs, is given by the ratio for which $\partial\widetilde{E(ht)}/\partial \ln GDPpc$ in Equation (2) for $X = \ln GDPpc$ becomes zero, i.e. by

$$\frac{\tilde{h}_2 - \tilde{h}_1}{\tilde{h}_3 - \tilde{h}_2} = \frac{\frac{\partial s_3}{\partial \ln GDPpc}}{\frac{\partial s_1}{\partial \ln GDPpc}}. \tag{3}$$

Thus, if the difference between "pretty happy" and "not too happy" is larger than the difference between "very happy" and "pretty happy" by a multiplicative factor equal to the ratio of the changes in the shares of "very happy" and "not too happy" respondents, we will observe a zero effect of a change in $\ln GDPpc$ on $\widetilde{E(ht)}$.[5] The coefficient estimates for $\ln GDPpc$ in the regressions of s_1 and s_3 in Table 1 imply that this multiplicative factor equals $0.0535/0.0254 \approx 2.11$. In the case of a three-points scale there is therefore a unique[6] transformed scale for reported happiness hr at which the sign of the effect of increases in $\ln GDPpc$ ($\partial\widetilde{E(ht)}/\partial \ln GDPpc$) switches. Hence, for hr scales that are more concave in the rank-order scale $(1, 2, 3)$ than such a scale, it is found that mean happiness in the USA rose rather than fell with increasing log *per capita* GDP (see Bond and Lang, 2019, Fig. A-2).

However, there turns out not to exist any hr scale which is concave enough to yield a statistically significant positive coefficient of $\ln GDPpc$ at the 5% or even 10% level. The best we can get is a positive coefficient of $\ln GDPpc$ with $p = 0.13$ in the limit for $\tilde{h}_3 - \tilde{h}_2 \to 0$, entailing $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2) \to \infty$. In that limiting case of an infinitely strongly concave hr scale (i.e. $(1, 2, 2)$; see Table 1), mean happiness $\widetilde{\widetilde{E(ht)}}$ coincides with $-s_1$, i.e. minus the share of the "not too happy". Consequently, the coefficient of $\ln GDPpc$ in the regression of $\widetilde{\widetilde{E(ht)}}$ is identical to minus the coefficient in the regression of s_1 (with the same $p = 0.13$). Thus, although there is a reported happiness scale at which the sign of the effect of $\ln GDPpc$ on mean happiness switches, there does not exist a sufficiently concave scale at which this sign reversal becomes significant at the 5% or 10% level. In that sense, the Easterlin Paradox for the USA cannot be rejected with any scale of reported happiness. This result is opposite to that obtained by Bond and Lang (2019), who found such significant sign reversals for sufficiently skewed latent happiness scales in an ordered probit model (Section A.3.1). This discrepancy of results is mainly due to Bond and Lang's use of a method different from ours which allows for variation in happiness with $\ln GDPpc$ within each response category. Their method will be investigated in Section 3. Finally, because the negative coefficient of $\ln GDPpc$ in the regression of $E(ht)$ for the linear rank-order scale of hr is not significant, the last column of Table 1 also reports estimates for the case of a reported happiness

[4] This higher weighting of the share of the "not too happy" should be distinguished sharply from the normative higher weighting of the happiness of the "not too happy" in a Rawlsian type of social welfare function. The former higher weighting is purely due to a non-normative concave scale of the underlying true happiness intensity ht.

[5] Note that Equation (2) also becomes zero if $\tilde{h}_3 - \tilde{h}_2 = \tilde{h}_2 - \tilde{h}_1 = 0$, but such a degenerate scale is excluded by the obvious constraint $\tilde{h}_1 < \tilde{h}_2 < \tilde{h}_3$.

[6] I.e. unique up to a positive linear transformation of the hr scale. See also footnote 2.

scale which is just convex enough to yield a significant negative coefficient of $\ln GDPpc$ (at the 5% level; found numerically).

2.2 Fundamental cause of possibility of sign reversal

In the example of the previous section, the possibility of a sign reversal of the effect of log *per capita* GDP on mean happiness $E(ht)$ is caused by the fact that the effects of log *per capita* GDP on both the share of the "very happy" (s_3) and that of the "not too happy" (s_1) are negative. Hence, while an increase in *per capita* GDP made some "not too happy" people "pretty happy" (i.e. happier), it also made some "very happy" people "pretty happy" (i.e. less happy).[7] Thus, the sign of the effect of *per capita* GDP on individual happiness is heterogeneous across its distribution. This is the fundamental cause of the possibility of sign reversals. Unfortunately, this point is not sufficiently recognized or emphasized in the theoretical analyses of Schröder and Yitzaki (2017) and Bond and Lang (2019). In the next section, we show that this point holds more generally for any number of response categories.

In the present example, we also observe that the cumulative response share of "not too happy" and "pretty happy" ($s_1 + s_2 = 1 - s_3$) increases with log *per capita* GDP ($\partial(s_1 + s_2)/\partial \ln GDPpc = -\partial s_3/\partial \ln GDPpc > 0$). At the same time, the cumulative response share of "not too happy" s_1 decreases with log *per capita* GDP. Hence, while $s_1 + s_2$ in a group of respondents with low log *per capita* GDP in a given year is lower than $s_1 + s_2$ in a group with high *per capita* GDP, s_1 in the group with low *per capita* GDP is higher than s_1 in the group with high *per capita* GDP. This means that neither cumulative distribution function of happiness responses for either low or high *per capita* GDP first-order stochastically dominates the other distribution. This violation of first-order stochastic dominance (FOSD) in the cumulative categories implies that we cannot conclude that a year group with low *per capita* GDP is happier than a year group with high *per capita* GDP, or the reverse, under all cardinal codings of reported happiness hr (cf. Schröder and Yitzaki, 2017, Condition 1, and Bond and Lang, 2019, Section 2).

Thus, the possibility of sign reversals of the effect of log *per capita* GDP on mean happiness $E(ht)$ is a symptom of an underling deficiency of the mean happiness model. This deficiency is the sign heterogeneity of the effect of *per capita* GDP on individual happiness, and hence on the cumulative response shares, across the happiness distribution. Such a sign heterogeneity indicates that the estimation equations for the (cumulative) response shares and mean happiness should be extended so as to account for the heterogeneity. One way to do this is to add control variables that are correlated with log *per capita* GDP to these equations. The omitted-variable bias in the estimates of the effect of log *per capita* GDP on the response shares may be heterogeneous across the happiness categories. Adding control variables may then diminish the sign heterogeneity of the effect of log *per capita* GDP on the response shares, and consequently the likelihood of sign reversal of the effect of log *per capita* GDP on mean happiness. Another way to extend the estimation equations so as to account for the sign heterogeneity is to add higher-order terms of log *per capita* GDP or interaction terms with other variables. Both ways to extend the estimation equations will be investigated in Section 5.1.

[7] As a result, the happiness inequality in the USA fell with increasing *per capita* GDP. Clark, Flèche, and Senik (2014; 2016) found that this is a standard pattern which holds for a large number of developed countries.

However, even if such extensions of the mean happiness model do not lead to a substantial decline in the likelihood of sign reversal of the effect of log *per capita* GDP on mean happiness, the question remains how plausible the skewed happiness scales that are required for sign reversals are. Intuitively, a value of 2.11 for the ratio $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2)$ that leads to a zero effect of log *per capita* GDP on mean happiness, may be plausible. However, according to our analysis there is no hr scale for which the sign reversal becomes significant at the 5% or 10% level.

2.3 General relabelling condition for sign reversal in mean happiness

In cases of happiness scales with more than three response options, the analysis becomes more complicated. Consider how mean happiness in a population varies with a certain explanatory variable X. Again, assume that the rank-order scale $(1, 2, \dots, R)$ for reported happiness hr is a good approximation of the average cardinal values of true happiness intensity ht within the R response categories.[8] Thus, within each category, the sample mean $E(ht|hr = k)$ equals k for $k = 1, 2, 3$. Mean happiness in the population is then estimated as the overall sample mean $E(ht) = s_1 * 1 + s_2 * 2 + \dots + s_R * R$, where s_k denotes the share of the sample that reports $hr = k$. Hence, when $E(ht)$ is linearly related to X, the effect of a unit change in X on $E(ht)$ is given by $\partial E(ht)/\partial X = 1 * \partial s_1/\partial X + 2 * \partial s_2/\partial X + \dots + R * \partial s_R/\partial X$. Writing the shares s_k as differences in successive cumulative shares $\sum_{l=1}^{k} s_l - \sum_{l=1}^{k-1} s_l$, we can rewrite the expression for $\partial E(ht)/\partial X$ as

$$\frac{\partial E(ht)}{\partial X} = (1-2) * \frac{\partial s_1}{\partial X} + (2-3) * \frac{\partial (s_1 + s_{2)}}{\partial X} + \dots + (R-1-R) * \frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X}$$

$$= -\frac{\partial s_1}{\partial X} - \frac{\partial (s_1 + s_{2)}}{\partial X} - \dots - \frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X}. \qquad (4)$$

Some of the derivatives of the cumulative shares may be negative while others may be positive. Suppose now that the negative derivatives dominate the positive derivatives. This would make $\partial E(ht)/\partial X$ in Equation (4) positive, implying that happiness rises with increasing X.

However, the rank-order scale $(1, 2, \dots, R)$ for reported happiness hr may not correctly reflect the differences in true happiness intensity ht between the R response categories. If so, $E(ht|hr = k)$ would not equal k for $k = 1, 2, \dots, R$. Denoting $E(ht|hr = k)$ as h_k for $k = 1, 2, \dots, R$, Equation (4) then becomes

$$\frac{\partial E(ht)}{\partial X} = (h_1 - h_2)\frac{\partial s_1}{\partial X} + (h_2 - h_3)\frac{\partial (s_1 + s_2)}{\partial X} + \dots + (h_{R-1} - h_R)\frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X}, \qquad (5)$$

where $E(ht) = \sum_{l=1}^{R} s_l * h_l$. We assume that $E(ht|hr = k)$ is monotonically increasing in k. Therefore, only labelling schemes for hr that satisfy $h_1 < \dots < h_R$ are allowed. Thus, all the "coefficients" $h_k - h_{k+1}$ of $\partial \sum_{l=1}^{k} s_l/\partial X$ are negative.

When all cumulative share derivatives are negative, the cumulative distribution function of happiness responses for high X stochastically dominates that for low X. In such a case the effect of X on mean happiness will be positive for any permitted labelling scheme of hr. However, when

[8] Some trivial adjustments to the argument below are necessary in order to cover the more general case in which the rank-order scale is a good approximation for average ht up to a linear transformation. We here omit these adjustments to aid notational simplicity.

at least one cumulative share derivative in Equation (5) is positive, there exist permitted labelling schemes for hr for which the effect of X on mean happiness will be negative. If most cumulative share derivatives are nevertheless negative, the differences $h_k - h_{k+1}$ between adjacent happiness responses for which the cumulative share derivatives are in fact positive will have to be (much) larger in size than the differences for negative cumulative share derivatives in order to achieve a reversal. *Mutatis mutandis*, the same lines of arguments hold for cases when most cumulative share derivatives are positive, but at least one derivative is negative. The results of this analysis can be summarised in the following proposition:

Proposition 1. The effect of an explanatory variable X on mean happiness $E(ht)$ cannot change sign by relabelling the scores of reported happiness hr if and only if all effects of X on the cumulative response shares have the same sign.

The hr scales beyond which the effect of X on mean happiness switches sign are given by the scales for which $\partial E(ht)/\partial X$ in Equation (5) becomes zero. When not all cumulative share derivatives in Equation (5) have the same sign (which can only occur for $R > 2$), there are infinitely many of such scales. For hr scales with three response options ($R = 3$), the previous section has shown that scales beyond which a sign reversal occurs, have a unique ratio $(h_3 - h_2)/(h_2 - h_1)$, which according to Equation (5) for $\partial E(ht)/\partial X = 0$ is given by $-(\partial s_1/\partial X)/(\partial(s_1 + s_2)/\partial X)$. For hr scales with more than three response options matters are more complicated. To see this, note that Equation (5) for $\partial E(ht)/\partial X = 0$ can be rewritten as

$$\frac{\partial s_1}{\partial X} + \frac{h_3 - h_2}{h_2 - h_1}\frac{\partial(s_1 + s_2)}{\partial X} + \frac{h_4 - h_3}{h_2 - h_1}\frac{\partial(s_1 + s_2 + s_3)}{\partial X} + \cdots + \frac{h_R - h_{R-1}}{h_2 - h_1}\frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X} = 0. \quad (6)$$

For $R > 3$ this is an equation in more than one unknown ratio variable, which does not have a unique solution. Hence, in order to get a unique solution, we have to impose a constraint on the ratio variables. One regular solution is obtained as follows. Assume that each subsequent true happiness difference $h_k - h_{k+1}$ in Equation (5) equals the same constant factor $r > 0$ times the preceding true happiness difference $h_{k-1} - h_k$. The hr scale is then multiplicative like the exponential scale of Bond and Lang (2019), where the constant factor r is e^c (see Section 3.1 for more details). The subsequent ratios $(h_{k+1} - h_k)/(h_2 - h_1)$ in Equation (6) can then be rewritten as powers r^{k-2}, turning Equation (6) into an $R - 2^{\text{th}}$ order polynomial equation in r

$$\frac{\partial s_1}{\partial X} + \frac{\partial(s_1 + s_2)}{\partial X}r + \frac{\partial(s_1 + s_2 + s_3)}{\partial X}r^2 + \cdots + \frac{\partial \sum_{l=1}^{R-1} s_l}{\partial X}r^{R-2} = 0. \quad (7)$$

According to Descartes' Rule of Signs[9], the number of positive real roots r^* of this equation is either equal to the number of sign switches of the cumulative share derivatives in Equation (7) from one derivative to the next, or less than that by an even number. Hence, when there is only one sign switch, there exists one unique positive real root r^* of Equation (7). This appears to be the prevalent case in happiness empirics (see the example in Section 2.1). However, when the number of sign switches of successive cumulative share derivatives is even (e.g. two), there may

[9] See https://en.wikipedia.org/wiki/Descartes%27_rule_of_signs.

not exist a positive real root.[10] On the other hand, when there is an odd number of sign switches, there exists at least one positive real root. The number of possible sign switches can never be larger than the order $R-2$ of the polynomial in Equation (7). So, if the number of response options R is three (the case in Section 2.1), only one sign switch is possible, but for more than three response options there can be more than one sign switch.

When a positive real root of Equation (7) exists, this can be solved analytically for up to fourth-order polynomials, but in general only numerically for higher-than-fourth-order polynomials.[11] However, for the case of only one sign switch in the derivatives in Equation (7), Bond and Lang (2019) have developed an alternative method to derive a simple analytical expression for a unique positive real root like r^*. This method will be explained in Section 3.1.

2.4 General relabelling condition for sign reversal in individual happiness

A limitation of the reversal condition in Equation (5) for $\partial E(ht)/\partial X = 0$ is that it is cumbersome to apply it to (almost) continuous variables X like household income. This requires computing cumulative shares for a large number of different values of X in the empirical sample. A simpler procedure runs as follows. Replace mean happiness $E(ht)$ by individual reported happiness hr_i and assume that hr_i is related to individual true happiness ht_i as $hr_i = ht_i + \eta_i$, where η_i is measurement error with mean zero. This implies $ht_i = hr_i - \eta_i$, and hence $E(ht_i|hr_i = h_k) = h_k$, i.e. the scale $(h_1, \dots, h_R)$ of reported happiness hr_i gives a good approximation of the average cardinal values of underlying true happiness ht_i for $k = 1, \dots, R$ (as also assumed in Section 2.1).[12] Usually, scale $(h_1, \dots, h_R)$ is coded to have equal intervals like the rank-order scale $(1, \dots, R)$. Next, assume that ht_i is linearly related to X_i and a vector of control variables $\boldsymbol{Z}_i$ as $ht_i = \alpha + \beta X_i + \boldsymbol{\gamma}'\boldsymbol{Z}_i + \varepsilon t_i$, $i = 1, \dots, N$, where error εt_i has a zero mean and is uncorrelated to X_i and $\boldsymbol{Z}_i$. This implies the linear OLS regression model

$$hr_i = \alpha + \beta X_i + \boldsymbol{\gamma}'\boldsymbol{Z}_i + \varepsilon r_i, \;\; i = 1, \dots, N. \tag{8}$$

Here error $\varepsilon r_i = \varepsilon t_i + \eta_i$ is discrete at given values of X_i and $\boldsymbol{Z}_i$ (like hr_i) and is supposed to have a zero mean and to be uncorrelated to X_i and $\boldsymbol{Z}_i$ ($E(\varepsilon r_i X_i) = 0$ and $E(\varepsilon r_i \boldsymbol{Z}_i) = 0$).[13] Moreover, for the purpose of statistical inference, we take into account that OLS estimates of parameters β and $\boldsymbol{\gamma}$ in Equation (8) are asymptotically normally distributed (Angrist and Pischke, 2009, Section 3.1.3). This allows the use of t and F-tests in sufficiently large samples. Note that these parameter estimates measure the effects of changes in X_i and $\boldsymbol{Z}_i$ on underlying true happiness ht_i.

---

[10] However, this does not mean that no sign reversals are possible in general. It just means that no reversals using this particular constraint may be found. In general, when there is at least one sign switch of the cumulative share derivatives in Equation (5), we can always find a permitted root of that equation by imposing an alternative constraint on the $(h_1, \dots, h_R)$ scale of hr. This constraint is letting the "coefficients" $h_k - h_{k+1}$ on all positive cumulative share derivatives in Equation (5) be -1 and letting the coefficients on all negative derivatives be some constant c. Denoting the sum of all positive derivatives as Δ^+ and the sum of all negative derivatives as Δ^-, we can easily solve for c from Equation (5) for $\partial E(ht)/\partial X = 0$ as $c = \Delta^+/\Delta^-$. This will always yield a negative, and hence permitted value of c. Such a reversal condition can be seen as a generalization of condition (3) for more than three response categories. However, because of its implausibility, the implied irregular scale does not appear empirically relevant.

[11] See: https://math.stackexchange.com/questions/200617/how-to-solve-an-nth-degree-polynomial-equation.

[12] In the context of this individual happiness model, the operator E now refers to population means.

[13] We do not assume a zero conditional mean ($E(\varepsilon r_i|X_i, \boldsymbol{Z}_i) = 0$) because we follow the "regression as a linear approximation approach" of Angrist and Pischke (2009; footnote 9 and p. 48).

Next, replace the cumulative shares by a set of dummies. Let $hd_{k,i}$ (**h**appiness **d**ummy) equal 1 when reported happiness hr_i of individual i is lower than or equal to h_k, for $k = 1, \dots, R-1$, and 0 otherwise. Furthermore, replace the identity $E(ht) = (h_1 - h_2)s_1 + (h_2 - h_3)(s_1 + s_2) + \cdots + (h_{R-1} - h_R)\sum_{l=1}^{R-1} s_l + h_R$ by the analogous identity[14]

$$hr_i = (h_1 - h_2)hd_{1,i} + (h_2 - h_3)hd_{2,i} + \cdots + (h_{R-1} - h_R)hd_{R-1,i} + h_R. \qquad (9)$$

Then, regress the happiness dummies on X_i and $\boldsymbol{Z}_i$ by OLS as

$$\begin{aligned} hd_{1,i} &= \alpha_{d1} + \beta_{d1}X_i + \boldsymbol{\gamma}_{d1}'\boldsymbol{Z}_i + \varepsilon d_{1,i}, \\ hd_{2,i} &= \alpha_{d1} + \beta_{d2}X_i + \boldsymbol{\gamma}_{d2}'\boldsymbol{Z}_i + \varepsilon d_{2,i}, \\ &\dots \\ hd_{R-1,i} &= \alpha_{dR-1} + \beta_{dR-1}X_i + \boldsymbol{\gamma}_{R-11}'\boldsymbol{Z}_i + \varepsilon d_{R-1,i}. \end{aligned} \qquad (10)$$

Here the errors $\varepsilon d_{k,i}$ are assumed to have a zero mean and to be uncorrelated to X_i and $\boldsymbol{Z}_i$. Note that $E(hd_{k,i}|X_i = X, \boldsymbol{Z}_i = \boldsymbol{Z}) = E(\sum_{j=1}^{N} hd_{k,j}/N \,|X_j = X, \boldsymbol{Z}_j = Z\ \forall j) = E(\sum_{l=1}^{k} s_l \,|X_j = X, \boldsymbol{Z}_j = Z\ \forall j)$. Therefore, our estimated predictions $\widehat{hd}_{k,i}$ are predictions of the cumulative response shares up to category k given X and $\boldsymbol{Z}$. Thus, $\partial\widehat{hd}_{k,i}/\partial X_i = \hat{\beta}_{dk} = \partial\,\widehat{\sum_{l=1}^{k} s_l}/\partial X_i$. Hence, if $\hat{\beta}_{d1}, \hat{\beta}_{d2}, \dots, \hat{\beta}_{dR-1}$ all have the same sign, no sign reversals of $\partial E(ht)/\partial X$ in Equation (5) are predicted to be possible by virtue of Proposition 1.

Furthermore, Equation (9) implies that predictions of hr_i as given by $\widehat{hr}_i = (h_1 - h_2)\widehat{hd}_{1,i} + (h_2 - h_3)\widehat{hd}_{2,i} + \cdots + (h_{R-1} - h_R)\widehat{hd}_{R-1,i} + h_R$ are equal to those obtained via a direct OLS regression of hr_i (with $h_1 = 1, \dots, h_R = R$, or any transform thereof, using any other scheme for h_k). Put differently, estimate $\hat{\beta}$ from the regression $hr_i = \alpha + \beta X_i + \boldsymbol{\gamma}'\boldsymbol{Z}_i + \varepsilon_i$, where hr_i may have been constructed using any labelling scheme for h_k, can be expressed in terms of the coefficient estimates of $hd_{k,i}$ as

$$\hat{\beta} = (h_1 - h_2)\hat{\beta}_{d1} + (h_2 - h_3)\hat{\beta}_{d2} + \cdots + (h_{R-1} - h_R)\hat{\beta}_{dR-1}. \qquad (11)$$

Therefore, the set of regressions of $hd_{k,i}$ can generate a sign reversal condition for the OLS estimate $\hat{\beta}$. We can then formulate:

Relabelling Condition. Estimate $\hat{\beta}$ of the effect of an explanatory variable X_i on reported happiness hr_i does not change sign by relabelling the scores of hr_i if and only if all estimates $\hat{\beta}_{dk}$ of the effects of X_i on the happiness dummies $hd_{k,i}$ in Equations (10) have the same sign.

We can apply Equations (8)-(11) for individual happiness to the Easterlin Paradox example given in Section 2.1. When imposing equal macro weights on each yearly wave of the GSS, the estimates are equivalent to those presented for mean happiness in Table 1 (cf. Angrist & Pischke, 2009, Section 3.1.2). For example, a regression of $hd_{1,i}$ on $\ln GDPpc$ yields the same coefficient as a

[14] Note that $\sum_{l=1}^{R} s_l = 1$ and that the identity for $E(ht)$ underlies Equation (5). Further note that $hd_{R,i} = 1$. Identity (9) can be easily seen to hold by noting that if $hr_i = h_k$, $hd_{1,i} = hd_{2,i} = \cdots = hd_{k-1,i} = 0$, $hd_{k,i} = hd_{k+1,i} = \cdots = hd_{R-1,i} = 1$, and hence the right-hand side of Equation (9) boils down to h_k.

regression of s_1 on $\ln GDPpc$, and a regression of $hd_{2,i}$ yields the same coefficient as minus the coefficient on $\ln GDPpc$ from a regression of s_3 (since $s_1 + s_2 = 1 - s_3$). Moreover, the scale $(\tilde{h}_1, \tilde{h}_2, \tilde{h}_3)$ of hr_i for which a regression of hr_i on $\ln GDPpc$ produces a zero coefficient, follows from Equation (11) for $X_i = \ln GDPpc$ and has the same ratio $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2) = -\hat{\beta}_{d1}/\hat{\beta}_{d2} = 2.11$ as the ratio for $\widetilde{\mathrm{E}(ht)}$ in Table 1 (see Appendix A for more details).

For the special case where the regression model in Equation (8) does not include control variables $\boldsymbol{Z}_i$, Schröder and Yitzhaki (2017) derive a sufficient condition for the possibility of sign reversal of the effect of a change in X_i on hr_i (as indicated by parameter β). However, this condition (Condition 2 in Schröder and Yitzhaki) is complicated and not very transparent. Moreover, Schröder and Yitzhaki do not present a systematic method to find a transformation of the rank-order scale of hr_i that reverses the sign of the regression coefficient of an explanatory variable X_i (the transformed scales that they present in Table A6 of their paper are irregular and rather ad hoc).[15]

2.5 Bounds on trade-off ratios of coefficients

When assessing the substantive implications of estimates, we may be more interested in ratios of coefficients than in their magnitude. For example, the estimation of shadow prices (Bertram and Rehdanz 2015; Levinson 2012; Luechinger 2009) or equivalence scales (Biewen and Juhasz 2017; Borah, Keldenich, and Knabe 2019; Rojas 2007) principally relies on ratios of coefficients. Unfortunately, when effects are not perfectly homogenous across the distribution of hr_i, ratios of coefficients are affected by transformations of hr_i even when no sign reversals are possible.

Fortunately, bounds on the ratio of two coefficients for any transformation of hr_i can be given. Let $\hat{\beta}$ and $\hat{\gamma}$ be coefficient estimates from a regression of a particular coding of hr_i on respectively X_i and Z_i (plus possibly other controls). Using Equation (11) in Section 2.4, we can write the ratio of these coefficient estimates as the ratio of sums of corresponding coefficient estimates resulting from regressions of $hd_{k,i}$:

$$\frac{\hat{\beta}}{\hat{\gamma}} = \frac{\sum_{k=1}^{R-1}(h_{kj} - h_{kj+1})\hat{\beta}_{dk}}{\sum_{k=1}^{R-1}(h_{kj} - h_{kj+1})\hat{\gamma}_{dk}}. \tag{12}$$

From this we can deduce:

1. When all ratios $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$ of coefficient estimates from the regressions of $hd_{k,i}$ take the same value ρ, Equation (12) reduces to $\hat{\beta}/\hat{\gamma} = \rho$. In that case, ratio $\hat{\beta}/\hat{\gamma}$ of coefficient estimates

[15] Schröder and Yitzaki (2017) also derive a necessary and sufficient condition for an unambiguous ranking of mean happiness in two groups. Translated in our terminology, this Condition 1 states that expected happiness in group A is higher than that in group B for all scales of hr_i if and only if the cumulative distribution function of underlying true happiness ht_i in group A first-order stochastically dominates that in group B. However, according to our analysis in Section 2.2, the weaker first-order stochastic dominance (FOSD) in cumulative response categories is enough for guaranteeing an unambiguous ranking of mean happiness in two groups. In our view, Schröder and Yitzaki make a mistake in their proof of Condition 1 in Appendix A5 of their paper by incorrectly assuming that hr_i as a function of ht_i ($h(s)$ in their notation) is differentiable. However, this cannot be true because $hr_i(ht_i)$ is a discontinuous step function. Taking this into account can easily be seen to imply an analogous condition in terms of FOSD in expected cumulative response categories.

from the regression of hr_i does not depend on the particular coding scheme $(h_1, h_2, \dots, h_R)$ of hr_i. The ratio is then invariant under any transformation of such a scheme.[16]

2. In general, we can view ratio $\hat{\beta}/\hat{\gamma}$ as a weighted average of all ratios $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$. By recoding hr_i such that $h_l - h_{l+1} = 0$ for all $l \neq k$ and $h_k - h_{k+1} < 0$, we can assign all the weight to one particular ratio $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$. Doing so we obtain $\hat{\beta}/\hat{\gamma} = \hat{\beta}_{dk}/\hat{\gamma}_{dk}$. This scale is not allowed because of the zero intervals for $l \neq k$, but it acts as a bound for the allowed scales with non-zero intervals. It then follows that for all permitted transformations of the coding scheme of hr_i ratio $\hat{\beta}/\hat{\gamma}$ is larger (smaller) than the smallest (largest) ratio among all $\hat{\beta}_{dk}/\hat{\gamma}_{dk}$. We can hence give lower and upper bounds for ratio $\hat{\beta}/\hat{\gamma}$.

3 Analysis of Bond and Lang's approach

3.1 Variation within response categories and sign reversals in the ordered probit model

Consider again the Easterlin Paradox example in Section 2.1. Equation (2) for the effect of $X = \ln GDPpc$ on mean happiness $\widetilde{E(ht)}$ implicitly assumes that the conditional expectations $E(ht|hr = k) = h_k, k = 1, 2, \dots, R$, do not vary with X. However, in this example the empirical finding that both the share of the "not too happy" and the share of the "very happy" fell with increasing $\ln GDPpc$ in the USA suggests that within the response category of the "not too happy" the rise in $\ln GDPpc$ may have led to less unhappiness and within the response category of the "very happy" to less happiness. This implies a rise in $E(ht|hr = 1) = h_1$ and a decline in $E(ht|hr = 3) = h_3$ with increasing $\ln GDPpc$, and hence would modify Equation (2) into a more flexible equation as given in Appendix B.

That equation turns out to have two implications (see Appendix B for more details): First, it is ambiguous whether, compared to the estimates of Table 1, allowing for variation within response categories increases or decreases the required ratio $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2)$ at which the effect of $lnGDPpc$ on mean happiness becomes zero. Using the method of Bond and Lang (2019), which allows for variation within response categories in an ordered probit model (see below), we find a somewhat lower value of the comparable ratio e^c of $e^{0.67} = 1.95$. Second, it becomes more likely that the effect of $\ln GDPpc$ on mean happiness is significantly positive in the limit for limit for $\tilde{h}_3 - \tilde{h}_2 \to 0$ (see the scale for $\widetilde{\widetilde{\mathrm{E}(ht)}}$ in Table 1). This may explain why Bond and Lang (2019) found significant sign reversals for finite values of the scale ratio e^c (9.49 at 5% significance level and 6.49 at 10% level).[17]

A limitation of this approach of allowing for variation within response categories is that the direction and magnitude of such variation are not observed in the empirical data. A partial way out of this is to postulate an ordered probit model of hr_i in terms of X_i and $\boldsymbol{Z}_i$. This is the approach of Bond and Lang (2019). A difference in this approach with our approach discussed in Section 2.4 and the approach of Schröder and Yitzhaki (2017) is that Bond and Lang (B&L) frame their

[16] Further see Van Praag & Ferrer-i-Carbonell (2008), Ch. 2 on the invariance of trade-off ratios of coefficients when effects are homogenous across the distribution of hr_i.

[17] In the case of more than three response categories, the analysis becomes less clear-cut because it then becomes much more ambiguous what the signs of $\partial(\tilde{h}_k - \tilde{h}_{k-1})/\partial X$ in the extended equation (see Appendix B) could plausibly be.

analysis in terms of mean expected happiness $E(ht_i|X_i, \boldsymbol{Z}_i)$ rather than individual happiness ht_i. Their proxy for true happiness ht_i in an ordered probit model of hr_i is latent happiness hl_i as defined by the linear equation[18]

$$hl_i = \alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l{}' \boldsymbol{Z}_i + \varepsilon l_i, \qquad i = 1, \dots, N. \tag{13}$$

Here, error εl_i is assumed to be normally[19] distributed with mean zero and standard deviation σ_i. B&L thus assume εl_i to be continuous and unbounded. A certain level of hl_i is reported as belonging to category k of hr_i for $k = 1, \dots, R$ if and only if $\tau_{k-1} \leq hl_i < \tau_k$. Here the τ_k's are cutoff points which are assumed to be common to all respondents, and $\tau_0 = -\infty$ and $\tau_R = \infty$.

In essence, this model is designed to estimate the signs of the effects of the explanatory variables X_i and $\boldsymbol{Z}_i$ on hl_i on the basis of merely ordinal information in the reported categories of hr_i (see, e.g., Ferrer-i-Carbonell and Frijters, 2004, Section 3.2, and Vendrik and Woltjer, 2007, Section 3.1). Implicitly, however, it also implies a cardinalisation of hl_i as defined by model equation (13) and the estimated set of τ_k's (see Van Praag and Ferrer-i-Carbonell, 2008, Section 2.5). B&L use this cardinalisation to compare mean happiness in two groups A and B with X_i in Equation (13) being a dummy variable D_i for group membership of A ($D_i = 1$) or B ($D_i = 0$). They maintain that the assumption in this model of a single cardinalisation under which both groups' happiness is distributed normally, is very strong. Further, they argue that the standard ordered probit assumption in happiness research of a constant variance $\sigma_i^2 = \sigma^2$ of εl_i across all groups is implausible and unnecessary for estimation, but necessary to obtain an unambiguous ranking of mean happiness in group A versus group B.[20] Moreover, they observe that, in large samples, variances across different groups will never be estimated to be *exactly* equal.[21] Therefore, neither group's cumulative distribution function (CDF) of hl_i will first-order stochastically dominate the other one. Hence, there will always exist alternative cardinalisations for which the ranking of mean happiness between groups A and B under the normal cardinalisation will be reversed.

In the context of Equation (13), mean happiness $E(ht_i|X_i, \boldsymbol{Z}_i)$ is given by $E(hl_i|X_i, \boldsymbol{Z}_i) = \alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l{}' \boldsymbol{Z}_i$. The effect of X_i on $E(hl_i|X_i, \boldsymbol{Z}_i)$ is then assessed by estimating parameter β_l in regression equation (13) for hl_i. Furthermore, the variance of εl_i is allowed to vary across persons by letting the standard deviation σ_i depend on X_i and $\boldsymbol{Z}_i$. The model $\ln(\sigma_i) = \alpha_s + \beta_s X_i + \boldsymbol{\gamma}_s{}' \boldsymbol{Z}_i$ is particularly appealing since σ_i can only take on values larger than 0 in this case. This model can then be estimated jointly with Equation (13) and the remaining set of cutoff points τ_k by maximum likelihood (see, e.g., Williams 2010). For identification, B&L further impose $\tau_1 = 0$ and $\tau_2 = 1$.

This joint estimation yields predictions of mean happiness $E(hl_i|X_i, \boldsymbol{Z}_i) = \alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l{}' \boldsymbol{Z}_i$ and (mean) variance $E({\sigma_i}^2|X_i, \boldsymbol{Z}_i) = {\sigma_i}^2 = e^{2(\alpha_s + \beta_s X_i + \boldsymbol{\gamma}_s{}' \boldsymbol{Z}_i)}$ together with estimates of the cutoffs τ_k. The scale of hl_i is given by these cutoff estimates. However, any positive monotone

[18] We add the notation of hl_i (rather than writing ht_i directly) since, as will be made clear below, hl_i as defined by Equation (13) is only one of many possible latent proxies of ht_i.

[19] Error εl_i could also be assumed to be logistically distributed, implying an ordered logit model. All the arguments to follow can be adapted to this model as well.

[20] They also note that this argument holds for any unbounded distribution from the location-scale family, not just for the normal distribution.

[21] The test statistics in heteroscedasticity tests have continuous distributions, and hence the a priori probability that these statistics are exactly zero, indicating perfect homoscedasticity, is zero.

transformation of this scale would yield the same likelihood. Thus, nothing in the data tells us which scale is most appropriate. One simple convex transformation of hl_i is e^{chl_i}, where c is a positive constant. This yields a transformed latent happiness variable $\widetilde{hl}_i$ with an exponential cutoff scale $\left(\tilde{\tau}_0, \tilde{\tau}_1, \ldots, \tilde{\tau}_R\right) = (0, 1, e^c, e^{c\tau_3}, \ldots, e^{c\tau_{R-1}}, \infty)$. This scale is similar to, but somewhat different from the multiplicative interval scale of reported happiness hr_i with constant ratio $r = e^c > 1$ that we have investigated in Section 2.3. The convex transformation changes the standard model (13) for hl_i into a new model for $\widetilde{hl}_i$:

$$\widetilde{hl}_i = e^{chl_i} = e^{c(\alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l' \boldsymbol{Z}_i + \varepsilon l_i)} \leftrightarrow$$

$$\frac{\ln \widetilde{hl}_i}{c} = hl_i = \alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l' \boldsymbol{Z}_i + \varepsilon l_i. \tag{14}$$

This model for transformed latent happiness $\widetilde{hl}_i$ is *a priori* not less plausible than model (13) for latent happiness hl_i. For example, if the explanatory variable X_i is the log of household income, model (13) for $\beta_l < 1/c$ implies, just as model (13), diminishing marginal utility of household income.

As is well-known[22], Equation (13) implies that the distribution of $\widetilde{hl}_i$ is log normal with mean

$$\tilde{\mu}_i = e^{c\mu_i + 0.5c^2 {\sigma_i}^2}, \tag{15}$$

where we use the short-hand notation μ_i for $E(hl_i | X_i, \boldsymbol{Z}_i)$. Since this expression for $\tilde{\mu}_i$ is increasing in ${\sigma_i}^2$, it follows that if μ_i rises with X_i, but ${\sigma_i}^2$ falls with X_i, the effect of X_i on $\tilde{\mu}_i$ will change sign and become negative for sufficiently large c. The value of c for which the effect of X_i on $\tilde{\mu}_i$ is predicted to become zero, and hence beyond which the effect should become negative, is easily derived by differentiating the expression in Equation (15) with respect to X_i and setting the derivative to zero. This yields

$$c = -\frac{2\frac{\partial \mu_i}{\partial X_i}}{\frac{\partial {\sigma_i}^2}{\partial X_i}} = -\frac{2\frac{\partial(\alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l' \boldsymbol{Z}_i)}{\partial X_i}}{\frac{\partial e^{2(\alpha_s + \beta_s X_i + \boldsymbol{\gamma}_s' \boldsymbol{Z}_i)}}{\partial X_i}} = -\frac{\beta_l}{e^{2(\alpha_s + \beta_s X_i + \boldsymbol{\gamma}_s' \boldsymbol{Z}_i)} \beta_s}. \tag{16}$$

When β_l and β_s have opposite signs, Equation (16) implies that the predicted sign-reversing value of c is positive. However, this value depends on the level of X_i as well as those of the control variables $\boldsymbol{Z}_i$. In our empirical applications, we report the sign-reversing value of c at the means of X_i and $\boldsymbol{Z}_i$. Since this value is positive, the multiplicative ratio e^c is larger than one, implying that the differences in happiness intensity between the cutoffs of successive happiness categories tend to increase from low to high levels of happiness.[23] The log-normal distribution of transformed reported happiness $\widetilde{hl}_i$ is then right-skewed.

[22] See, e.g., https://en.wikipedia.org/wiki/Log-normal_distribution.

[23] The difference in $\widetilde{hl}_i$ between cutoffs $\tilde{\tau}_k$ and $\tilde{\tau}_{k-1}$ differs from the difference between cutoffs $\tilde{\tau}_{k-1}$ and $\tilde{\tau}_{k-2}$ by a factor $(e^{c\tau_k} - e^{c\tau_{k-1}})/(e^{c\tau_{k-1}} - e^{c\tau_{k-2}})$. Since the cutoff differences $\tau_k - \tau_{k-1}$ and $\tau_{k-1} - \tau_{k-2}$ under the normal cardinalisation tend to be rather similar to each other (say equal to $\Delta\tau$, except for the extreme cutoffs), this factor turns out to be approximately equal to $e^{c\Delta\tau}$, which is larger than one.

Alternatively, if both μ_i and ${\sigma_i}^2$ rise (or fall) with X_i (implying that β_l and β_s have the same sign), the sign-reversing value of c in Equation (16) is negative. The positive sign of the effect of X_i on μ_i is then reversed by concave rather than convex transformations of hl_i, such as $-e^{chl_i}$ with $c < 0$. This yields a transformed latent happiness variable $\widetilde{hl}_i$ with an exponential cutoff scale $(\tilde{\tau}_0, \tilde{\tau}_2, \dots, \tilde{\tau}_R) = (-\infty, -1, -e^c, -e^{c\tau_3}, \dots, -e^{c\tau_{R-1}}, 0)$. This scale is again similar to, but somewhat different from the multiplicative interval scale of hr_i with constant ratio $r = e^c < 1$. Differences in happiness intensity between the cutoffs of successive happiness categories then tend to decrease from low to high levels of happiness. In this case, the log-normal distribution of transformed reported happiness $\widetilde{hl}_i$ is left-skewed with mean

$$\tilde{\mu}_i = -e^{c\mu_i + 0.5c^2{\sigma_i}^2}, \tag{17}$$

which is decreasing in ${\sigma_i}^2$. Thus, a sufficiently strongly negative c reverses the positive sign of the effect of X_i on $\tilde{\mu}_i$. The value of c beyond which this is predicted to happen is again given by Equation (16), but now it is negative.

B&L do not derive condition (16), but instead present a special case of it for the comparison of mean happiness in two groups A and B. This case corresponds to $X_i = D_i$ in condition (16), which is then given by $c = -2\Delta\mu_i/\Delta{\sigma_i}^2 = 2(\mu_A - \mu_B)/(\sigma_B^2 - \sigma_A^2)$. However, B&L nevertheless illustrate their approach with a large number of examples for the more general case of more than two groups as identified by explanatory variable X_i (see the Online Empirical Appendix to their paper). Their first example is the Easterlin Paradox example that we have analysed in Sections 2.1 and 2.4. On the same data as used in these sections, we estimate a heteroskedastic ordered probit (HOP) model as described in Equation (13) and with $\tau_1 = 0$ and $\tau_2 = 1$. Estimation of this model, which is largely equivalent to Bond & Lang's model[24], yields estimates -0.045 and -0.165 for the marginal effects of $X_i = \ln GDPpc$ on μ_i and $\ln(\sigma_i)$, respectively. The former estimate is very close to coefficient -0.043 of $\ln GDPpc$ in column (2) of Table A-1 of B&L (2019)[25], but our coefficient is just not significant ($p = 0.13$) because of our clustering of the standard errors by year (which B&L unfortunately omitted). Applying condition (16) at the mean of $\ln GDPpc$ to our coefficients yields a c value of -0.73 for which the effect of $X_i = \ln GDPpc$ on transformed mean happiness $\tilde{\mu}_i$ is predicted to become zero. This value is close to the value -0.67 that we find in a search procedure such as used by B&L.[26] Using the same numerical search, one can also find the

[24] We have programmed this model in a somewhat different and more straightforward way than Bond and Lang. This produces identical results. The Stata do file is available on request from the authors.

[25] The difference in size of the coefficient is due to our use of equal macro weights for each yearly wave of the GSS.

[26] For their search procedure, B&L use a more flexible variant of their HOP model in which μ_i and σ_i are estimated separately for each year. This is equivalent to a HOP model with year dummies instead of $X_i = \ln GDPpc$ in Equation (13) and the linear regression equation for $\ln(\sigma_i)$ (cf. p. A-25 of B&L (2019)). The slight discrepancy in c values between estimate -0.67 from this search procedure and the predicted -0.73 from condition (16) at mean $\ln GDPpc$ is mainly due to the strong non-linearity of $\ln(\sigma_i)$ in $\ln GDPpc$. When running a numerical search on the basis of the HOP model in which $\ln GDPpc$ is entered linearly, we find a value of -0.73 which agrees with the predicted value. More generally, the last expression in condition (16) predicts the sign-reversing level of c to lie in a range between the value of the expression for the highest level of $\ln GDPpc$ ($= 10.80$; yielding $c = -0.81$) and the value for the lowest level of $\ln GDPpc$ ($= 10.13$; yielding $c = -0.65$) in the sample. This range includes the c value of -0.67 from B&L's search procedure.

level of c that is required to make the effect of $\ln GDPpc$ on transformed mean happiness $\tilde{\mu}_i$ just significantly positive at the 5% level ($c = -4.34$).[27]

B&L (2019) present their parametric analysis only in the context of an ordered probit model. However, by estimating Equation (8) together with the equation $\ln \sigma_i = \alpha_s + \beta_s X_i + \boldsymbol{\gamma_s}' \boldsymbol{Z}_i$ by maximum likelihood (see, e.g., Gould, Pitblado, and Poi 2010), their approach can in principle also be applied to the discrete interval model for hr_i discussed in Section 2.4. Such an application would rest on the assumption that $\varepsilon r_i = \varepsilon t_i + \eta_i$ in Equation (8) is normally distributed (at given values of X_i and $\boldsymbol{Z}_i$). This assumption is obviously incorrect because it implies that error εr_i is continuous and unbounded whereas it actually is discrete and bounded (at given X_i and $\boldsymbol{Z}_i$). When nevertheless carrying out such an estimation, we obtain estimates of -0.029 and -0.082 for the effects of $\ln GDPpc$ on hr_i and $\ln(\sigma_i)$, respectively. Applying condition (15) at the mean of $\ln GDPpc$ to these estimates and a search procedure such as used in the HOP model above yield sign-reversing c values of -1.77 and -1.61, which are much more negative than what we found for the HOP model. Accordingly, $c = -1.61$ implies a ratio $r = e^{-1.61} = 0.20$ of the multiplicative scale of $\widetilde{hr_i}$ which is much smaller than the ratio $0.47 = 1/2.11$ found in Table 1 and Section 2.4. This strong underestimation of the sign-reversing scale ratio is likely to be due to the misspecification of the distribution of error εr_i in Equation (8) as normal.

3.2 Comparison of models

The two most commonly used models in the happiness literature are the OLS discrete interval model and the ordered probit model. Nowadays, the discrete interval model is more widely used than the ordered probit model because it typically yields similar results but is much easier to estimate and interpret. For the interval model, the sign reversal analysis given in Section 2.4 is clearly more appropriate than the analysis summarized at the end of the previous section. An important result of Section 2.4 is that a sign reversal of the effect of X_i on hr_i by relabelling the scores of reported happiness scales is only possible if our relabelling condition is violated. As we will see in Section 5 and in contrast to B&L's approach, this condition is easily satisfied in many empirical applications, implying that sign reversals are impossible in these cases. However, the ordered probit model of B&L implies that transforming cardinal happiness scales by transforming underlying identifying models practically always allows for a sign reversal. This raises the question where this difference in results comes from.

The analysis in Section 3.1 suggests that a crucial difference between B&L's and our approach is that B&L's ordered probit model allows for variation of latent happiness hl_i with X_i and $\boldsymbol{Z}_i$ within response categories whereas our discrete interval model for hr_i does not. This implies that in B&L's analysis first-order stochastic dominance (FOSD) in cumulative distribution functions of hl_i of one group relative to another one is required to exclude sign reversals (see B&L (2019), Section 2). Our approach only requires the weaker property of FOSD in the cumulative response categories of hr_i to rule out sign reversals (see Section 2.2). This is an important difference because FOSD in

[27] Bond and Lang obtain $c = -2.25$, because they do not use standard errors which are robust to autocorrelation and heteroscedasticity. Although a Breusch-Pagan test for heteroskedasticity does not reject homoscedasticity at any reasonable significance level, Durbin's alternative test for autocorrelation yields a rejection of the null hypothesis of no autocorrelation with $p < 0.05$ for a lag order of up to 2. We therefore use Newey-West standard errors that are robust to such autocorrelation.

cumulative distribution functions from the location-scale family will never be satisfied in large samples because variances across different groups will never be estimated to be exactly equal. In contrast, FOSD in cumulative response categories is equivalent to the relabelling condition mentioned above, and hence holds in many empirical cases.

We can get a deeper understanding of the differences in approach by considering the first line of Equation (14) in the B&L model for $\widetilde{hl}_i$ more closely. An analogous expression applies to our approach. Using a relabelling of the sort described in Section 2.3, we may obtain $\widetilde{hr}_i = e^{chr_i}$. Our approach thus implies a transformation of the linear model given by Equation (8) in Section 2.4. However, this transformation is not made explicit because $\widetilde{hr}_i$ is *directly* linearly regressed on X_i and $\boldsymbol{Z}_i$, i.e. we estimate $\widetilde{hr}_i = \tilde{\alpha} + \tilde{\beta}X_i + \tilde{\gamma}\boldsymbol{Z}_i + \tilde{\varepsilon r}_i$. Error $\tilde{\varepsilon r}_i$ in this regression is again discrete and bounded at given values of X_i and $\boldsymbol{Z}_i$.[28] In contrast, in B&L's "*indirect*" approach, first $\tilde{\mu}_i \equiv E(\widetilde{hl}_i|X_i, \boldsymbol{Z}_i)$ is derived as given by Equation (15), and then, $\tilde{\mu}_i$ is linearly regressed on X_i and $\boldsymbol{Z}_i$. The sign reversal condition for this regression is given by Equation (16). This condition can be derived in an alternative way which gives a deeper insight in the similarities as well as essential differences between B&L's and our approach.

First, write the normally distributed error εl_i in Equation (13) as $\sigma_i\epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0,1)$. The derivative $\partial\mu_i/\partial X_i \equiv \partial E(hl_i|X_i, \boldsymbol{Z}_i)/\partial X_i$ in condition (16) can then be rewritten as

$$\frac{\partial E(hl_i|X_i, \boldsymbol{Z}_i)}{\partial X_i} = \frac{\partial \int_{-\infty}^{\infty} hl_i(\epsilon_i)\,\varphi(\epsilon_i)d\epsilon_i}{\partial X_i} = \int_{-\infty}^{\infty} \frac{\partial hl_i}{\partial X_i}(\epsilon_i)\,\varphi(\epsilon_i)d\epsilon_i = E\left(\frac{\partial hl_i}{\partial X_i}\middle| X_i, \boldsymbol{Z}_i\right).^{29} \quad (18)$$

Here $\varphi(\epsilon_i)$ is the standard normal density function $(2\pi)^{-0.5}e^{-0.5\epsilon_i{}^2}$ and $hl_i(\epsilon_i)$ indicates that hl_i is a function of the integration variable ϵ_i (besides X_i and $\boldsymbol{Z}_i$). Thus, the (marginal) effect of X_i on mean happiness equals the mean effect of X_i on individual happiness. Equation (13) and the relation $\sigma_i = e^{\ln\sigma_i} = e^{\alpha_s+\beta_s X_i+\boldsymbol{\gamma_s}'\boldsymbol{Z}_i}$ imply that the "local" effect $\partial hl_i/\partial X_i$ of a unit change in X_i on hl_i for a given value of error ϵ_i equals $\beta_l + \beta_s\sigma_i\epsilon_i$. The sign of this local effect depends on the value of ϵ_i. For example, suppose that effect β_l of X_i on μ_i is positive and that effect β_s of X_i on $\ln\sigma_i$ is negative. Then the local effect of X_i on hl_i will be negative (i.e. $\beta_l + \beta_s\sigma_i\epsilon_i < 0$) for a sufficiently large value of ϵ_i. Clearly, this is the case when $\epsilon_i > -\beta_l/(\beta_s\sigma_i) > 0$. This implies that the sign of the local effect of X_i on hl_i is heterogeneous across the distribution of ϵ_i, and hence of hl_i. In Section 2.2 such heterogeneity has been shown to represent the fundamental cause of the possibility of sign reversal of the effect of an explanatory variable X_i on mean happiness.

For untransformed hl_i, Equation (18) implies that the resulting overall mean effect of X_i on hl_i equals $\beta_l + \beta_s\sigma_i E(\epsilon_i|X_i, \boldsymbol{Z}_i) = \beta_l$, and hence is positive in the present example. However, for a sufficiently strong transformation $\widetilde{hl}_i = e^{chl_i}$, the sign of the mean effect of X_i on $\widetilde{hl}_i$ may be different. This effect again equals the effect of X_i on transformed mean happiness $\tilde{\mu}$, and it can be

[28] However, error $\tilde{\varepsilon r}_i$ has of course different values than εr_i. More specifically, $\tilde{\varepsilon r}_i$ is given by $\tilde{\varepsilon r}_i = \widetilde{hr}_i - \tilde{\alpha} - \tilde{\beta}X_i - \tilde{\gamma}\boldsymbol{Z}_i = \pm e^{chr_i} - \tilde{\alpha} - \tilde{\beta}X_i - \tilde{\gamma}\boldsymbol{Z}_i$.

[29] This uses the mathematical property that the derivative of the integral with respect to X_i in the second term of this equation can be brought inside the integral because the integration limits and the integration variable ϵ_i do not depend on X_i.

expressed as an integral over local effects on $\tilde{\mu}_i$. To see this, differentiate the first line of Equation (14) with respect to X_i and integrate the resulting expression over the standard normal distribution as

$$E\left(\frac{\partial \widetilde{hl}_i}{\partial X_i}\middle| X_i, \boldsymbol{Z}_i\right) = \pm c e^{c(\alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l' \boldsymbol{Z}_i)} \int_{-\infty}^{\infty} e^{c\sigma_i \epsilon_i}(\beta_l + \beta_s \sigma_i \epsilon_i)\varphi(\epsilon_i) d\epsilon_i. \qquad (19)$$

The integrand in this equation again changes sign beyond $\epsilon_i = -\beta_l/(\beta_s \sigma_i)$. However, now the weights $e^{c\sigma_i \epsilon_i}\varphi(\epsilon_i)$ of the local effects $\beta_l + \beta_s \sigma_i \epsilon_i$ in the overall mean effect of X_i on $\widetilde{hl}_i$ are relatively higher for large positive ($c > 0$) or large negative ($c < 0$) values of ϵ_i as compared to the weights $\varphi(\epsilon_i)$ in the integrand in Equation (18). Hence, in the above example for $\beta_l > 0$ and $\beta_s < 0$, the negative local effects for $\epsilon_i > -\beta_l/(\beta_s \sigma_i) > 0$ in the right half of the standard normal distribution get higher weights for $c > 0$. These weights increase in c, implying that for sufficiently large c the negative local effects will start to dominate the positive local effects in making the overall mean effect of X_i on $\widetilde{hl}_i$, and hence the effect of X_i on $\tilde{\mu}$, negative. The value of c beyond which this will occur, is predicted by a reversal condition that follows from integrating the integral in Equation (19) by parts. Appendix C shows that this condition turns out to be condition (16).

Thus, just as relabelling condition (11) in our approach, reversal condition (16) in B&L's approach turns out to be essentially based on heterogeneity in sign of local effects of X_i on individual happiness. Note that in this case the heterogeneity is implied by heteroscedasticity of error εl_i in Equation (13).[30]

However, why is a sign reversal always possible in B&L's approach, but not so in our approach? On the one hand, in B&L's model the standard normal density function $\varphi(\epsilon_i)$ of error ϵ_i is unbounded. Therefore, even if ratio $-\beta_l/(\beta_s \sigma_i)$ beyond which the local effect of X_i on hl_i changes sign is very large positive or negative, there will always exist sufficiently extreme values of ϵ_i in a tail of its distribution for which the local effect of X_i on hl_i has a different sign. On the other hand, in our model (8) for reported happiness hr_i the distribution of error εr_i at given values of X_i and $\boldsymbol{Z}_i$ is inherently bounded. Hence, if ratio $-\beta_l/(\beta_s \sigma_i)$ is very large positive or negative, there will not always exist sufficiently extreme values of ϵ_i in one of the tails of the distribution of εr_i for which the local effect of X_i on hr_i switches sign.[31]

A key issue in judging whether B&L's approach or our approach is more appropriate in evaluating whether sign reversals are possible, is whether the underlying scale of true happiness ht_i is bounded or not. Introspection does not give us a clear answer to this question. On the one hand, there may be upper and lower limits to one's happiness as expressed by the labels "completely happy" and "completely unhappy". On the other hand, for every happiness level one can imagine, one may be able to imagine another happiness level at which one is even happier or even less happy. However,

[30] This is because heterogeneity of local effects of X_i on hl_i is equivalent to a conditional expectation function $E(hl_i|X_i, \boldsymbol{Z}_i)$ that is nonlinear in X_i (cf. footnote 13). Hence, if one uses a linear regression model to approximate such a nonlinear conditional expectation function, it reveals itself as heteroscedasticity of error εl_i (see Angrist and Pischke, 2009, p. 46).

[31] Strictly speaking, this argument presumes that we approximate the discrete distribution of εr_i at given X_i and $\boldsymbol{Z}_i$ by a continuous, but bounded underlying distribution of true happiness ht_i (e.g., ranging from 0.5 to $R + 0.5$ for the usual rank-order scale $(1, \ldots, R)$ of hr_i).

in practice the scale of ht_i is bounded by limited variation in human biology and finite numbers of people in populations.

A more empirically minded argument in favour of our approach is the following. Consider a situation in which all empirically observed cumulative response shares move in the same direction when an explanatory variable X_i rises (or falls). Our relabelling condition is then satisfied and sign reversals are not possible in our model. But in B&L's model they still are. The analysis above showed that such sign reversals are then made possible by sign reversals of local effects of X_i on individual happiness in an extreme tail of its distribution. However, such sign reversals of local effects are not empirically observed, and hence there is no clear indication in the data that they occur. Therefore, in such situations the possibility of sign reversals of the overall effect of X_i on mean happiness is highly speculative.

4 Plausibility of multiplicative happiness scales

4.1 General and theoretical arguments

When we transform a rank-order scale of reported happiness hr (or a similar scale of latent happiness hl in the ordered probit case) into a multiplicative scale $\widetilde{hr}$ (or similar scale $\widetilde{hl}$), and take such a scale as a better proxy for ht, we also change our substantive beliefs about the way in which persons reply to happiness questions. When hr is rank-order coded we believe that the difference in mean ht between each pair of adjacent response categories is constant. When we transform this scale, these differences are no longer constant. Instead, the difference in mean ht between levels of $\widetilde{hr}$ grows or declines by a multiplicative factor r which is larger or smaller than one, respectively (see Section 2.3). For response scales with just three categories this may not be too problematic. For instance, in order to just reverse the effect of per capita GDP in the Easterlin Paradox example in Sections 2.1 and 2.4, we only require a value of $r = e^{-0.75} = 0.47$ in our direct approach and a similar ratio $e^{-0.67} = 0.51$ in B&L's HOP model. These ratios imply that a jump in true happiness intensity from the 2nd to the 3rd response category is roughly half as big as a jump from the 1st to the 2nd response category. It seems possible that respondents use the response scale in this manner. However, scales that lead to a 5% significant reversal do not even exist for the effect of $\ln GDPpc$ on $\widetilde{hr}$ in the EP example and seem much too extreme to be plausible for the effect of $\ln GDPpc$ on $\widetilde{hl}$ in the HOP model ($e^{-4.34} = 0.01$).

This plausibility problem becomes even more severe when the number of response categories grows large. For instance, the question on life satisfaction in the oft-used German SOEP survey has 11 response categories. When for example applying an exponential transformation e^{chr} with $c = 1$ (which is smaller than the just-sign-reversing c's in most of our relevant results in Section 5.2), the difference in ht between $hr = 0$ (the lowest coding) and $hr = 1$ becomes 1.72. However, the difference between $hr = 9$ and $hr = 10$ becomes 13,923.38, which is more than 8,000 times larger.[32] Thus, when applying such a transformation and treating our transformed variable $\widetilde{hr}$ as an equally plausible proxy of ht, we say that the following two ways of responding to a happiness question are equally plausible:

[32] $e^1 - e^0 = 1.72$ and $e^{10} - e^9 = 13923.38$.

1. Respondents interpret each difference between subsequent steps[33] on their response scale as covering equal distances on their scale of experienced ht. This may be illustrated as:

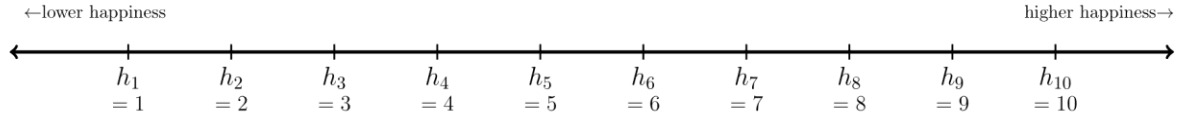

Figure 1. Linear response scale with equal intervals.

2. Respondents interpret each difference between subsequent steps on their response scale as covering a distance on their experienced scale of ht that is larger by a constant multiplicative factor $r = e^c$. This behavior is illustrated below for $r = e^1 = 2.72$:

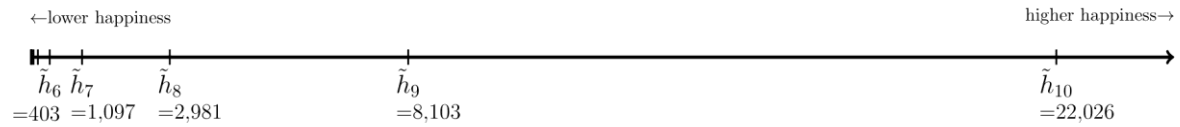

Figure 2. Multiplicative response scale with exponentially increasing intervals ($r = e^1 = 2.72$).

From introspection, it appears to us that the former interpretation is closer to our way of answering happiness questions. However, we may not be representative.

A more general, theoretical argument in favour of a linear response scale as illustrated by Figure 1 is provided by Van Praag (1971). Respondents may think of a finite set of response categories as discretizing an underlying continuous quantity (like true happiness). When reasoning about how to use such a set of categories to discretize that quantity, respondents may attempt to maximize the information that they give in the questionnaire. They can do so by minimizing the expected "inaccuracy" of their answer[34], as modelled with a cost function like the square of the prediction error (i.e. the expectation of the square of the difference between what a respondent feels and what a researcher infers from that answer). Van Praag shows that if respondents believe the underlying quantity to be uniformly distributed, discretizing this quantity into equally spaced intervals minimizes this cost function. Kapteyn (1977) generalizes this result to hold for any plausible cost function.

4.2 Experimental evidence of approximately linear response scales

As a first piece of experimental evidence, Van Praag (1991) tested how persons translate ordered verbal labels (*very bad; bad; not bad; not good; good; very good*) into cardinal quantities in a context-free setting. In a first experiment he asked subjects to assign numbers between 1 and 1000 to each verbal label. Here, 1 was said to stand for the "very worst" and 1000 for the "very best". In a second experiment, respondents were asked to produce lines of certain length corresponding to each verbal label. Here a length of 1 unit was said to stand for the "worst", and a length of 40 to stand for the "best". No further context was given in either experiment. Scaling both sets of responses to lie on the interval [0,1], Van Praag's results were as follows:

[33] Depending on whether one views the scale of ht as bounded or not, the top and bottom categories are exempted from that.

[34] See Bless, Strack, and Schwarz (1993) for arguments in favour of that assumption, and Van Praag (1991) and Parducci (1995) for related efficiency arguments.

Table 2. Van Praag's results

	Numbers experiment		Lines experiment	
Verbal label	Mean (90% CI)	Interval	Mean (90% CI)	Interval
(1) *Very bad*	0.089 (0.081-0.097)		0.073 (0.068-0.078)	
(2) *Bad*	0.201 (0.191-0.212)	0.112	0.180 (0.172-0.188)	0.107
(3) *Not bad, not good*	0.472 (0.462-0.482)	0.271	0.401 (0.392-0.410)	0.221
(4) *Good*	0.668 (0.658-0.678)	0.196	0.598 (0.588-0.608)	0.197
(5) *Very good*	0.866 (0.857-0.874)	0.198	0.823 (0.813-0.833)	0.225

Note: Confidence intervals are computed using a student's t-distribution with sample standard deviations given on p. 78 of van Praag (1991) and N = 361.

Both experiments suggest roughly linear scale use. Although there is some variation in intervals between verbal labels, there is no obvious pattern to it. At most, it appears that labels "very bad" and "bad" are less strongly distinguished than all the other labels. Van Praag explains this by noting that respondents may be trying to "leave room" for values associated with the other more positive labels.

As more direct evidence, Studer (2012) analyses Dutch data in which respondents were asked to report their happiness with a slider on a continuous and bounded scale.[35] Using such a scale, respondents are enabled to report their ht more directly. This makes it plausible to assume that values obtained from such a slider can be interpreted cardinally. Fortunately, respondents were also asked to report their happiness on a more standard Likert scale with ten response options (ranging from 0 to 9).[36] Studer can therefore directly compare response behaviour from a question that can reasonably be assumed to measure ht cardinally (up to a linear transformation), and a question that measures hr in a manner that is common in typical happiness research. In doing so, he evaluates which partitioning of the continuous scale would reproduce the observed response shares on the discrete Likert scale. More formally, he finds the set of cutoffs $\tau_k, k = 0, \dots, 10$, that satisfy $\tau_0 = 0$ and $F_{cont.}(\tau_k) = \sum_{l=0}^{k-1} s_l$, where $F_{cont.}$ is the empirical CDF of responses for the question using the continuous slider and s_l denotes the share of respondents that report response category l on the question using the Likert scale (as in Section 2.4). He finds that differences between cutoffs do not form a regular pattern: they are neither equidistant nor do they follow a multiplicatively increasing or decreasing scale. However, ratios of almost all subsequent differences of cutoffs do not exceed 1.8 or fall below 0.4.[37] The only exceptions are the ratios $(\tau_2 - \tau_1)/(\tau_1 - \tau_0)$ and $(\tau_3 - \tau_2)/(\tau_2 - \tau_1)$. These ratios are more extreme since very few people responded with the lowest two categories on the Likert scale. Studer's results therefore suggest that scales with multiplicative factors $r < 0.4$ and $r > 1.8$, corresponding to $c < -0.92$ and $c > 0.59$ are not plausible for Likert scales with relatively many response categories.

Finally, an additional argument in favour of a linear response scale can be derived from experimental results in psychophysics. Because this argument is rather intricate we relegate it to Appendix D. Thus, we have suggestive evidence from various experiments and theoretical

[35] The boundedness of the scale may be justified by assuming that happiness, as argued in Section 3.2, is bounded between states of being completely happy (or satisfied) and being completely unhappy (or dissatisfied).

[36] In both cases, the original question reads "Alles bij elkaar genomen, hoe gelukkig zou u zeggen dat u bent?" (Taking all things together, how happy would you say you are?), with extremes labelled "helemaal ongelukkig" (completely unhappy) and "helemaal gelukkig" (completely happy).

[37] That is to say: $0.4 \leq (\tau_k - \tau_{k-1})/(\tau_{k-1} - \tau_{k-2}) < 1.8 \; \forall \; k \in 4 \dots, 9$. Unfortunately, Studer (2012) only provides a graphical figure and not precise cutoff values. These are therefore estimates.

perspectives that all point in a similar direction: People are most likely to use roughly linear response scales.[38]

4.3 Scales with few response categories may be multiplicative

However, it is possible that the preceding arguments only hold in cases with at least five response categories.[39] The happiness variables that are mostly analysed by B&L have just three or four response categories. We may think that respondents use these scales as collapsed versions of scales with more categories. If so, multiplicative scale use may be plausible in such cases. We therefore perform a similar exercise to that of Studer (2012).

Table 3 lists cumulative shares in each response category from the happiness question of the 2006 wave of the US General Social Survey (GSS; with three categories), as well as cumulative shares for a life satisfaction question from the 5th (2006) wave of the United States sample in the World Values survey (WVS).[40] Both samples are representative of the same population and the two questions measure strongly correlated concepts of ht.

Table 3. Cumulative response shares for happiness and life satisfaction in GSS and WVS

GSS		WVS		Mean hr
hr	Share in % (cum.)	hr	Share in % (cum.)	after collapse
		1 ("Completely dissatisfied")	0.46 (0.46)	
		2	0.90 (1.36)	
1 ("Not too happy")	11.98 (11.98)	3	2.05 (3.41)	4.14
		4	3.89 (7.30)	
		5	7.32 (14.61)	
		6	9.71 (24.32)	
2 ("Pretty happy")	55.80 (67.78)	7	23.06 (47.38)	7.30
		8	28.27 (75.65)	
		9	17.65 (93.29)	
3 ("Very happy")	32.22 (1.00)	10 ("Completely dissatisfied")	6.71 (100.00)	9.28

Note: Data from GSS and WVS wave 5 (both 2006). Design weights applied.

The observed cumulative response shares in these samples suggest that category "not too happy" in the GSS questions most closely corresponds to categories 1-5 on a 10-points scale. Likewise, category "pretty happy" seems most likely to correspond to categories 6-8 and category "very happy" corresponds to categories 9-10 on a 10-points scale. Assume now that the relative distribution of responses across the 10-points scale in the WVS sample (measuring life satisfaction) is a reasonable approximation of the distribution of responses we would observe had the GSS sample (measuring happiness) been given a 10-points scale. Assuming that the 10-points scale measures ht roughly cardinally (as argued in the previous section), we can then take mean hr across categories 1-5 of the WVS variable as indicative of mean ht in the "not too happy" response category of the GSS variable. This yields a mean of 4.14. Same arguments apply to mean hr of

[38] Another suggestive piece of evidence in favour of linearity comes from Oswald (2008). He shows that when people are asked to rate their own height on a bounded scale, they treat that scale as linear. This is shown by regressing responses on the bounded scale against true height and the square of true height. While the coefficient on the squared terms for true height is negative, its magnitude is rather small, implying only a negligible degree of concavity in the response scale. However, this result may be particular to heights. Other quantities may be rated in a non-linear manner. For example, subjective loudness of a sound is recorded on a logarithmic decibel scale of its physical intensity.

[39] This is the number of categories used in van Praag's experiments.

[40] Unfortunately, we are not aware of a publicly available dataset that has a 10-points or 11-points scale for a question on happiness in the United States.

categories 6-8 (mean = 7.30) and 9-10 (mean = 9.28) of WVS as being indicative of mean ht in categories "pretty happy" and "very happy" of GSS. These (assumed) differences in ht between response categories of the GSS variable become smaller by a ratio of $(9.28 - 7.30)/(7.30 - 4.14) = 0.62$.[41]

To just reverse the effect of per capita GDP in B&L's EP example, we required a ratio of 0.47 when using our direct method and a ratio of 0.51 when using B&L's method. Since these are not much more extreme than the ratio 0.62 just obtained, reversals for the EP example may be plausible. However, in order to obtain a reversal which is significant at the 5% level, we require a ratio of 0.01 when using B&Ls method (for our method no such scale exists). Such a scale seems very implausible given the discussion in the previous section.

Furthermore, using WVS (4-points scale for happiness) and ESS (11-points scale for happiness) data, we also applied a similar procedure to a set of 14 European countries. That exercise yielded that differences between responses on the 4-point WVS scale collapse in a roughly linear manner onto the 11-point ESS scale. See Table A1 of Appendix E for results.[42] It therefore appears that convex/concave scales of the degree B&L require (see, e.g., Section A3.4) may be plausible for questions with three response options, but less so for questions with more response options.

Finally, the argument of B&L in Sections 3 and A1.3 that it is plausible that the distribution of (true) happiness is more skewed than that of wealth and comparable to that of income is strange in view of empirical evidence of diminishing marginal happiness from income (Vendrik and Woltjer, 2007; Layard et al., 2008). This evidence implies that, if anything, the distribution of happiness (at given values of the explanatory variables) is considerably less right-skewed than the distribution of income or may even be left-skewed. In fact, the left-skewedness of the distribution of the continuous happiness variable in Fig. 4 of Studer (2012) strongly suggests that the distribution of true happiness may be left-skewed as well. This may also hold when we condition that distribution on a set of control variables because relatively little variation in happiness is typically explained by such controls.[43] However, we do not know of any argument why the degree of left-skewedness of such a distribution should be comparable to the degree of right-skewedness of the income or wealth distribution.

5 Empirical Applications

5.1 Adding control variables and using scales with many response categories

In Section 2.2 we speculated that the likelihood of reversals can be reduced by adding relevant controls, which may reduce the heterogeneity in the effect of $\ln GDPpc$ across the distribution of

[41] The matching of GSS happiness shares and WVS life satisfaction shares in Table 3 takes into account that people tend to be happier than they are satisfied with their life (see e.g. Knabe et al. 2010). This is confirmed by a comparison of shares of happiness and life satisfaction scores on 11-points scales in wave 3 (2006) of the ESS (see Table A1 in the Appendix).

[42] The figures for mean hr after collapse in the fifth column of Table A1 imply adjacent happiness differences from "not at all happy" to "very happy" of 2.71, 3.56, and 2.64. The subsequent ratios of these differences are given by 1.31 and 0.74. The corresponding c values of 0.27 and -0.30 are much smaller in size than the c values of 1.72 and -1.72 in Table A-3 of B&L between which country rankings in mean happiness reverse.

[43] An important exception are individual fixed effects as used in our empirical application in Section 5, but these are not included in the empirical examples of B&L (2019). Moreover, B&L use few control variables in most of these examples.

reported happiness. In the context of the Easterlin Paradox, a particularly salient control is a linear time trend that picks up secular trends in other determinants of mean happiness than $\ln GDPpc$ (see Kaiser & Vendrik, 2019). In Table A2 of Appendix E, we thus extend Table 1 by adding a linear time trend to the estimation equations. This causes the effect of $\ln GDPpc$ on the share of those responding "very happy" to become positive while the effect on the share of those "not too happy" remains negative, and the effect on those "rather happy" remains positive. Consequently, the effect of $\ln GDPpc$ on mean rank-order-coded hr becomes significantly positive (while the coefficient of year is significantly negative). The Easterlin Paradox is therefore rejected in this case[44], and no transformation of hr could reverse this sign of the effect of $\ln GDPpc$.[45]

We now turn to assessing the wider empirical relevance of the points of the preceding sections. We do so by evaluating the possibility and plausibility of reversals for a range of important demographic variables using waves 1 (1984) to 32 (2015) of the German Socio-Economic Panel (GSOEP). The GSOEP is a nationally representative survey of the German population and is among the most commonly used dataset in empirical happiness economics. Our explanatory variables of interest are household income, unemployment, marriage, having children, and self-reported disability. These variables are similar to those investigated in the empirical appendix of Bond and Lang.[46] Answers to the question *"How satisfied are you with your life, all things considered"*[47], with 11 response categories labelled from 0 to 10 are used as our dependent variable hr_i. Bond and Lang's analyses rely on questions with only three to seven response categories. Since 10 or 11 response categories are more typical for most happiness research, it will be useful to study whether plausible reversals can be obtained with the present variables. Moreover, the effects of income and unemployment have been extensively studied with these data. Consequently, there is now near universal agreement that, at least in the short run, higher incomes improve life satisfaction while unemployment reduces it. It would be a disturbing finding if these results were easily reversed.

For income we use log net (post-tax) household incomes, deflated to 2005 prices. We equivalize incomes using the modified OECD scale.[48] Regarding unemployment, we code a dummy that is 1 when a person reports to be unemployed, and 0 for any other possible employment status. We code similar dummies for being married, living with children in the household, and reporting a disability. Next to reporting results in which these variables are entered separately, we also report

[44] However, the positive coefficient of $\ln GDPpc$ is likely to pick up business cycle effects which should be controlled for in a proper test of the Easterlin Paradox (see Kaiser and Vendrik, 2019 for details).

[45] Adding further controls like ln(individual household income), unemployment status, the square of $\ln GDPpc$, and its interaction with ln(individual household income) or unemployment status does not change this result, although the effects of ln(individual household income), unemployment status, and their interactions with $\ln GDPpc$ on mean rank-order hr are significant.

[46] In their Appendix A.3.3, they consider the effects of the unemployment rate and inflation rate on happiness using Eurobarometer data. Unfortunately, they do not report which level of c would (just) reverse the sign of the effect of unemployment. In Section A.3.6, they consider the effect of being married and of having children using BHPS data. They find that a right-skewed transformation with $c = 0.32$ reverses the effect of being married for men, while a left-skewed transformation with $c = -2.69$ reverses the effect of being married for women. Left-skewed transformations with $c = -0.74$ and $c = -0.64$ reverse the effect of children in the household for men and women, respectively. Also using BHPS data, in Section A.3.8 Bond and Lang show that a right-skewed transformation with $c = 1.41$ would reverse the effect of disability. None of their analyses except that of the effects of unemployment and inflation include control variables.

[47] In German: „Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?" (DIW 2016).

[48] We exclude respondents in the top and bottom percentiles of the income distribution as well as the self-employed, since there may be substantial measurement error in these observations (Berthoud and Bryan 2011; Hurst, Li, and Pugsley 2013).

results in which all variables are entered jointly along with a set of additional control variables. The additional control variables include region ("Bundesländer") dummies, wave dummies, age, age squared, a tertiary education dummy, a home ownership dummy, log household size and log working hours.[49] We restrict our sample to those above the age of 18. In total, our sample includes 557,999 observations. We first present results for our method using relabelling, and then turn to Bond & Lang's method.

5.2 Reversals using relabelling

Ferrer-i-Carbonell & Frijters (2004) made evident that time-invariant unobserved heterogeneity - driven by e.g. individual personality traits (Boyce, 2010) - causes considerable bias in pooled-regression estimates. Applied researchers have consequently turned to individual-fixed-effects models to account for this unobserved heterogeneity. Unfortunately, no fixed-effects estimator is readily available for the ordered probit model. This is a major drawback to Bond & Lang's method.[50] However, by demeaning each regression of $hd_{k,it}$ (index t runs over time periods) our relabelling condition can be straightforwardly applied to the fixed-effects model.[51]

Thus, Table 4 shows results for pooled and fixed-effects regressions of hr_{it} on each of our explanatory variables of interest. Columns (1) and (3) show results from separate regressions in which each variable is entered individually (being married and having children are always entered jointly), while columns (2) and (4) show results from regressions in which all variables of interest, along with the additional controls discussed above, are entered jointly.[52] In all specifications, household income and being married are associated with higher life satisfaction, while unemployment and reporting a disability are associated with lower life satisfaction. Having children in the household is also associated with higher satisfaction but turns insignificant when including fixed effects as well as controls. More generally, accounting for fixed effects reduces the size of coefficient estimates for every explanatory variable.

In order to evaluate whether the sign of these coefficients can be reversed, we estimate regressions of $hd_{k,it}$ for $k = 0,1,\dots,9$ when entering variables separately, when including controls, and when either running pooled regressions or when controlling for fixed effects. Tables A3 and A4 in Appendix E show our full results. To illustrate, Figure 3 plots estimated coefficients of household income for each regression of $hd_{k,it}$ (see Figures A1 to A4 in Appendix F for such plots for unemployed, married, having children and disability). Recall that our relabelling condition states that coefficient sign reversals are impossible if and only if all coefficient estimates on $hd_{k,it}$ have the same sign for all k.

[49] Entered as log(1+working hours) to allow for observations with zero working hours.
[50] A similar drawback is faced in Schröder & Yitzhaki's (2017) approach since their reversal condition does not allow for controls at all. See, however, Ferrer-i-Carbonell & Frijters (2004), Section 4, for an ordered logit model with individual-fixed effects.
[51] In this case, with full controls, linear age drops out of the estimations since it is then perfectly collinear with the individual-fixed effects and the wave dummies.
[52] Since one may worry that marriage and having children mediate the effect of income, we also ran regressions in which these variables are excluded. This yielded very similar results for all other variables.

Table 4. OLS and fixed effects regressions

	Pooled		Fixed effects	
	(1)	(2)	(3)	(4)
	No controls	Full controls	No controls	Full controls
Log household income	0.691*** (0.011) reversing **c=3.18**	0.568*** (0.012) reversal impossible	0.228*** (0.011) reversing **c=1.44**	0.296*** (0.011) reversal impossible
Unemployed	-1.273*** (0.019) reversal impossible	-0.917*** (0.018) reversal impossible	-0.643*** (0.016) reversal impossible	-0.638*** (0.015) reversal impossible
Married	0.189*** (0.012) reversal impossible	0.290*** (0.013) reversal impossible	0.171*** (0.014) reversing **c=2.17**	0.168*** (0.014) reversal impossible
Children	0.175*** (0.012) reversal impossible	0.132*** (0.012) reversing **c=−2.83**	0.068*** (0.012) reversal impossible	0.008 (0.012) reversing **c=0.13**
Disability	-0.857*** (0.021) reversal impossible	-0.766*** (0.020) reversal impossible	-0.495*** (0.019) reversal impossible	-0.306*** (0.018) reversing **c=2.53**
Observations	557,999	557,999	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Columns (1) and (3) result from separate models for each explanatory variable (being married and having children entered jointly). Reversing c values have been obtained numerically.

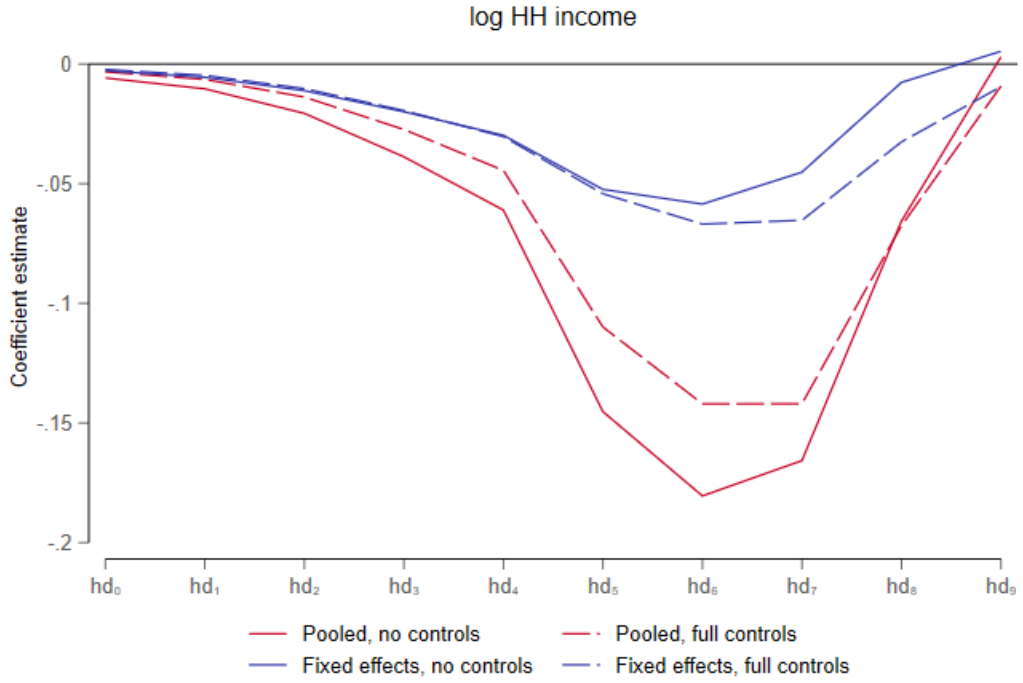

Figure 3. Coefficient estimates of income for each regression of $hd_{k,it}$, corresponding to Appendix Tables A3 and A4.

For our pooled regressions, we find that the income coefficient can be reversed when no controls are included. Since the sign of the effect on $hd_{9,it}$ is positive, while all others are negative, a sufficiently convex transformation can achieve such a reversal. A numerical search reveals that a multiplicative scale in which spaces between adjacent response categories grow by a ratio of 24.1 is required to just achieve such a reversal. This corresponds to a level of $c = \ln(24.1) = 3.18$ in the transformation $\widetilde{hr}_{it} = e^{chr_{it}}$. Second, the effect of having children can be reversed when including controls. Here, since the sign of the effect on $hd_{0,it}$ is positive while the effects in all other regressions of $hd_{k,it}$ are negative, a sufficiently concave transformation can achieve such a reversal. A numerical search yields that a transformation $\widetilde{hr}_{it} = -e^{chr_{it}}$ with $c = -2.83$ or lower is sufficient. Given the results of Table 4, these pooled estimates are likely biased due to the omission of fixed effects. Nevertheless, when accounting for fixed effects, reversals remain

possible for both income (without controls, convex, with $c = 1.44$ or higher) and for children (now convex with $c = 0.13$ or higher).[53] In addition, the effects of being married (without controls) and disability (with controls) can now be reversed by respective convex transformations with at least $c = 2.17$ and $c = 2.53$.

The effect of unemployment cannot be reversed in any specification. We may thus conclude that the common finding that unemployment is associated with lower life satisfaction is particularly robust. Should any of the other reversals worry us? In light of the arguments given in in Section 4, scales with c values in the order of at least 1 ($r \geq e^1 = 2.72$) or at most -1 ($r \leq e^{-1} = 0.37$) are implausible. Almost all just-sign-reversing scales found above fall within these ranges. The only exception is the reversal scale for $c = 0.13$, implying $r = e^{0.13} = 1.14$. This scale is reasonably close to a linear scale. We thus conclude that while reversals are *possible* for most variables in at least some specifications, the only *plausible* reversal is that of the effect of having children in a fixed-effects regression with controls. Since that result was strongly insignificant and close to zero in Table 4, this is not a particularly striking result.[54]

5.3 Bounds on trade-off ratios

However, Figures 3 and A1-A4 indicate that the relative magnitudes of the effects of each explanatory variable on $hd_{k,it}$ are not the same for all k. Given the discussion in Section 2.5, trade-off ratios of coefficients will consequently not be invariant under all transformations of hr_{it}. Section 2.5 also established that ratios of coefficients will be bounded from above and below by the largest and smallest corresponding ratios obtained from regressions of $hd_{k,it}$. Figure 4 plots the ratio of the coefficients of unemployment, being married, having children and disability against the coefficient for income in each of the fixed effects regressions of $hd_{k,it}$ (with full controls; corresponding to the bottom panel of Table A4).

Since the magnitudes of the coefficients of being married and having children are small, their ratios with the coefficients of income vary only little across regressions of $hd_{k,it}$. Moreover, since the coefficients of having children switch sign for $k \geq 7$, the ratios of these coefficients with those of income then change sign, too. For unemployment and disability, we observe that the ratios of these coefficients to the income coefficient generally increase with higher k. Therefore, the ratios of the effects of unemployment and disability on hr_{it} to the effect of income on hr_{it} will also increase (decrease) for increasingly convex (concave) transformations of hr_{it}.[55]

Using such ratios of coefficients, we can calculate the shadow price of each explanatory variable. We define the shadow price of e.g. unemployment as the amount of additional income needed for

[53] Similar to our findings for the Easterlin Paradox in which the addition of a linear time trend made reversals impossible (see section 5.1), further robustness regressions not shown here indicate that the inclusion of wave dummies is sufficient to make reversals of the sign of the coefficient of log household income impossible.

[54] Moreover, to achieve a *significantly* negative effect estimate with $p < 0.05$ of children we require $c = 0.65$, implying $r = e^{0.65} = 1.92$.

[55] Recall that convex (concave) transformations give relatively more weight to changes at higher (lower) levels of hr_{it}.

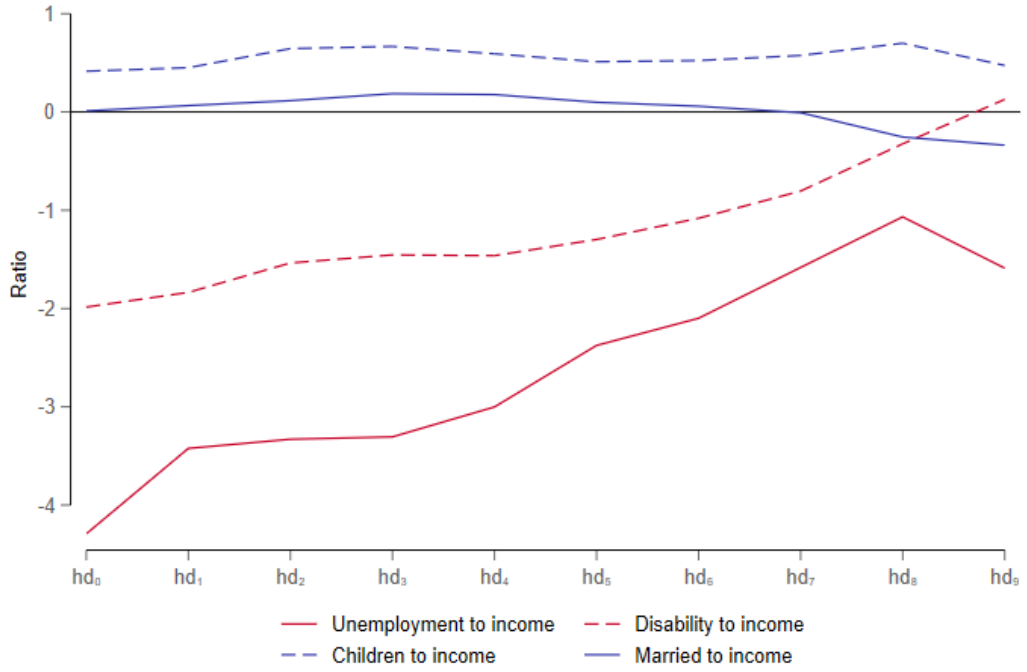

Figure 4. Ratio of coefficients of unemployment, being married, children and disability to income for each regression of $hd_{k,it}$. Regressions include the full set of controls and account for fixed effects (corresponding to the bottom panel of Table A4).

an unemployed person with a particular income level y to be as satisfied as someone who is not unemployed. This amount is given by $\left(e^{-\beta_{ue}/\beta_{lny}} - 1\right)y$.[56] Here, β_{lny} is the coefficient of log household income, and β_{ue} is the coefficient of unemployment. Consequently, the shadow price of unemployment will fall in a range determined by the largest and smallest ratio of coefficients obtained from regressions of $hd_{k,it}$. Ranges of shadow prices for being married, having children, and disability can be found analogously. In Table 5, we list ranges of shadow cost for each independent variable as evaluated at the mean of income.

Table 5. Shadow prices for each explanatory variable

Scale	Unemployment	Marriage	Children	Disability
Rank-order hr_{it}	€146,050 $(\frac{-0.638}{0.296})$	-€8,278 $(\frac{0.168}{0.296})$	-€500 $(\frac{0.008}{0.296})$	€34,677 $(\frac{-0.306}{0.296})$
Lower bound shadow price	€36,376 $(\frac{-0.035}{0.033})$	-€9,598 $(\frac{0.023}{0.033})$	€7,669 $(\frac{-0.003}{0.033})$	-€2,296 $(\frac{0.001}{0.009})$
Upper bound shadow price	€1,376,076 $(\frac{-0.010}{0.002})$	-€6,478 $(\frac{0.001}{0.002})$	-€3,249 $(\frac{0.004}{0.002})$	€119,769 $(\frac{-0.005}{0.002})$
$\widetilde{hr}_{it} = e^{chr_{it}}$, with $c = 0.4$	€92,782 $(\frac{-3.135}{1.772})$	-€8,370 $(\frac{1.024}{1.772})$	€1,108 $(\frac{-0.100}{1.772})$	€21,968 $(\frac{-1.358}{1.722})$
$\widetilde{hr}_{it} = e^{chr_{it}}$, with $c = -0.4$	€273,885 $(\frac{-0.037}{0.014})$	-€8,227 $(\frac{0.008}{0.014})$	-€1,590 $(\frac{0.001}{0.014})$	€53,100 $(\frac{-0.018}{0.014})$

Note: Calculations on the basis of Table 4, column (4), lower panel of Table A4 and fixed-effects regression of $\widetilde{hr}_{it}$ with $c = 0.4$ or $c = -0.4$ and full controls (not shown). Corresponding ratios of coefficients in parentheses. Negative shadow prices imply that a variable is estimated to benefit respondents. Thus, at the sample mean of household income and when using rank-order hr_{it}, a person who is *not* married needs to be compensated with €8,278€ of additional household income in order to be as satisfied as a person who is married.

As expected on the basis of Figure 4, we find that the ranges of estimated shadow prices of unemployment and disability cover an extremely wide range. In contrast, shadow prices for

[56] To see this, solve $\left[\beta_{lny}\ln(y + \Delta y) + \beta_{ue}\right] - \beta_{lny}\ln(y) = 0$ for Δy.

marriage and children are comparatively small and vary little.[57] However, these ranges of possible shadow prices rely on rather implausible transformations of hr_{it} in which differences between response categories approach zero except for some particular chosen response category (cf. Section 2.5). We therefore also evaluate how shadow prices change for a transformation $\widetilde{hr}_{it} = \pm e^{chr_{it}}$, with $c = 0.4$ and $c = -0.4$. These levels of c imply that differences in life satisfaction intensity between adjacent response categories increase or decrease by a factor of $e^{0.4} \approx 1.5$. We take it that such transformations may still be plausible.[58] This exercise shows that shadow prices for unemployment and disability still cover a rather wide range. Indeed, viewing these transformations as the most extreme plausible transformations, we do not know whether an unemployed (disabled) person can be compensated with as little as €93,000 (€22,000) or requires as much as €274,000 (€53,000).

We thus conclude that although sign reversals of the effects of explanatory variables on life satisfaction tend to be either impossible or very implausible, ratios of coefficients are heavily affected under even reasonably mild transformations. We therefore recommend that future empirical work should verify the robustness of its key results against mild convex and concave transformations of hr_{it}. We would also welcome for future work to at least check whether sign reversals of key coefficients are possible, and if so, to assess whether the required scale transformations seem plausible.

5.4 Reversals using Bond and Lang's method

We now turn to applying Bond and Lang's method to achieve reversals as discussed in Section 3.1. To do so, in Table 6 we report results from ordered probit regressions on the same data as in Table 4. We again observe that higher incomes, being married, and having children are always associated with higher (mean) life satisfaction while unemployment and disability are associated with lower satisfaction. The magnitudes of these coefficients are roughly twice those obtained in the pooled OLS results of Table 4. This is because differences between cutoffs are estimated to be somewhat above 2 for high levels of hr_{it} and somewhat below 2 for low levels of hr_{it}. Coefficients are therefore scaled by a factor of approximately 2 when compared to the rank-order coding used in Table 4. Concerning the estimated standard deviation in latent satisfaction, σ_{it} we find the opposite result: higher incomes, being married and having children reduce σ_{it}, while unemployment and disability increase σ_{it}. Since no coefficient on $\ln(\sigma_{it})$ is estimated to be precisely zero, reversals are possible for every variable.[59] Because every explanatory variable has different coefficient signs for μ_{it} as compared to $\ln(\sigma_{it})$, convex transformations with positive c will yield reversals.

[57] Note that since sign reversals were possible for children and disability, the signs of their shadow prices also depend on the chosen scale.

[58] The arguments of Section 4 only establish that extreme departures from linearity are not plausible. They do not establish exactly when departures from linearity become implausible, which is why these choices for c are of course somewhat arbitrary.

[59] However, the coefficient of children in the specification with full controls is insignificant and very close to zero. In turn, this small coefficient led to the extremely large estimate of the required value for c in Table 7.

Table 6. Heteroskedastic ordered probit regressions

	(1) HOP, variables entered separately	(2) HOP, full controls
$\boldsymbol{\mu_{it}}$		
Log HH income	1.453*** (0.043)	1.209*** (0.039)
Unemployed	-2.711*** (0.075)	-1.759*** (0.055)
Married	0.408*** (0.029)	0.577*** (0.031)
Children	0.409*** (0.030)	0.320*** (0.028)
Disability	-1.804*** (0.062)	-1.525*** (0.055)
Constant		10.336*** (0.224)
$\boldsymbol{\ln(\sigma_{it})}$		
Log HH income	-0.140*** (0.004)	-0.065*** (0.005)
Unemployed	0.066*** (0.006)	0.069*** (0.006)
Married	-0.038*** (0.004)	-0.049*** (0.005)
Children	-0.021*** (0.004)	-0.001 (0.005)
Disability	0.097*** (0.007)	0.073*** (0.007)
Constant		1.281*** (0.024)
Cutoff points		
τ_0		-∞ (assumed)
τ_1		0.000 (assumed)
τ_2		1.000 (assumed)
τ_3		2.377*** (0.036)
τ_4		3.775*** (0.068)
τ_5		4.875*** (0.094)
τ_6		7.032*** (0.145)
τ_7		8.366*** (0.177)
τ_8		10.499*** (0.228)
τ_9		13.781*** (0.306)
τ_{10}		16.257*** (0.365)
τ_{11}		∞ (assumed)
Observations	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Column 1 results from separate models for each explanatory variable (being married & having children entered jointly). Since constants and cutoff points vary (slightly) across regressions in column 1, they are not reported here.

Strikingly, the numerical c values in Table 7 show that reversals are not generally made harder with additional controls. Thus, in contrast with the Easterlin Paradox case in Section 5.1, adding such controls does not seem to reduce possible heterogeneities in the effects of explanatory variables across the distribution of hr_{it}. However, in order to give a correct interpretation of the degrees of skewedness of the transformed happiness scales that are implied by the computed c values in Table 7, we make the following observations. The large majority (about 75%) of the respondents in our sample reports happiness scores in the range of 5 to 9. In that range the differences between the estimated cutoff points in Table 6 do not deviate much from their average of 2.3. Therefore, for any given c, a transformation $\widetilde{hl}_{it} = e^{chl_{it}}$ is roughly 2.3 times as extreme as a similar transformation of rank-order coded hr_{it}. Hence, for a proper comparison with the latter transformations, we should multiply the c values in Table 7 by 2.3.

Table 7. Bond and Lang's reversal conditions

		log HH income	Unemployed	Married	Children	Disability
No controls	c (via Eq. 18)	0.66	2.33	0.61	1.14	1.06
	c (numerical)	0.66	2.34	0.60	1.15	1.06
Full controls	c (via Eq. 18)	1.43	1.84	0.90	26.08	1.53
	c (numerical)	0.68	0.69	0.51	1.55	1.59

Note: Analytical reversal condition (16) for income is evaluated at the sample means of all explanatory variables.

After this multiplication, it turns out that no numerical c value is below our benchmark of $c = 1$ from Section 4.1. The closest to this benchmark is the value required to just reverse the effect of being married with full controls, where we obtain $c = 2.3 * 0.51 = 1.17$. This implies an average ratio of $e^{1.17} = 3.22$ of subsequent differences between transformed cutoff points $\tilde{\tau}_k = e^{c\tau_k}$.

6 Conclusions

In this paper we showed that reversals are fundamentally caused by variables having heterogenous effects across the distribution of happiness. We derived necessary and sufficient conditions under which reversals of OLS regression coefficients are made feasible. We also derived bounds on trade-off ratios of coefficients of explanatory variables under any permissible labelling scheme. Moreover, we argued that in cases where reversals by "relabelling" are impossible, reversals using Bond & Lang's method are empirically unfounded. Furthermore, we presented arguments and evidence showing that respondents likely use the response scale in a roughly linear fashion. Finally, using GSOEP data, we empirically investigated the possibility and plausibility of reversals for a set of key variables. We found that reversals using relabelling are largely impossible or implausible and that reversals using Bond & Lang's indirect method are mostly implausible.

Thus, if the goal is to identify the direction of effects of explanatory variables, the worries flagged by Bond & Lang (2019) and Schröder & Yitzhaki (2017) do not appear to be a serious concern in practice. Nevertheless, since effects of explanatory variables are rarely homogenous across the distribution of happiness, trade-off ratios of coefficients may be severely affected by plausible transformations of reported happiness scores. We therefore urge researchers to verify the sensitivity of their results against plausible transformations of reported happiness. Moreover, we recommend that researchers ascertain whether our relabelling condition is satisfied for their application, thus verifying (or not) that their results are immune to reversals.

A limitation of our analysis is that we ignore any potential problems arising from heterogeneities in scale use. If such heterogeneities are correlated with explanatory variables, severe biases in estimated effects are possible. Hence, future work should carefully analyse these issues.

Lastly, our finding that the relative effects of explanatory variables are not homogenous across the distribution of happiness shows that estimating mean effects on happiness hides patterns in the data that are interesting and informative in their own right. As was previously done using quantile regressions (Binder and Coad 2011; 2015; Gupta et al. 2015), such patterns should be investigated more broadly.

References

Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press.

Banks, William P., and Mark J. Coleman. 1981. 'Two Subjective Scales of Number'. *Perception & Psychophysics* 29 (2): 95–105. https://doi.org/10.3758/BF03207272.

Banks, William P., and David K. Hill. 1974. 'The Apparent Magnitude of Number Scaled by Random Production.' *Journal of Experimental Psychology* 102 (2): 353.

Berthoud, Richard, and Mark Bryan. 2011. 'Income, Deprivation and Poverty: A Longitudinal Analysis'. *Journal of Social Policy* 40 (1): 135–156. https://doi.org/10.1017/S0047279410000504.

Bertram, Christine, and Katrin Rehdanz. 2015. 'The Role of Urban Green Space for Human Well-Being'. *Ecological Economics* 120: 139–52. https://doi.org/10.1016/j.ecolecon.2015.10.013.

Biewen, Martin, and Andos Juhasz. 2017. 'Direct Estimation of Equivalence Scales and More Evidence on Independence of Base'. *Oxford Bulletin of Economics and Statistics* 79 (5): 875–905. https://doi.org/10.1111/obes.12166.

Binder, Martin, and Alex Coad. 2011. 'From Average Joe's Happiness to Miserable Jane and Cheerful John: Using Quantile Regressions to Analyze the Full Subjective Well-Being Distribution'. *Journal of Economic Behavior & Organization* 79 (3): 275–90. https://doi.org/10.1016/j.jebo.2011.02.005.

———. 2015. 'Heterogeneity in the Relationship Between Unemployment and Subjective Wellbeing: A Quantile Approach'. *Economica* 82 (328): 865–91. https://doi.org/10.1111/ecca.12150.

Bless, Herbert, Fritz Strack, and Norbert Schwarz. 1993. 'The Informative Functions of Research Procedures: Bias and the Logic of Conversation'. *European Journal of Social Psychology* 23 (2): 149–165.

Bond, Timothy N., and Kevin Lang. 2019. 'The Sad Truth about Happiness Scales'. *Journal of Political Economy* 127 (4): 1629–40. https://doi.org/10.1086/701679.

Borah, Melanie, Carina Keldenich, and Andreas Knabe. 2019. 'Reference Income Effects in the Determination of Equivalence Scales Using Income Satisfaction Data'. *Review of Income and Wealth* 65 (4): 736–70. https://doi.org/10.1111/roiw.12386.

Clark, Andrew E., Sarah Flèche, and Claudia Senik. 2014. 'The Great Happiness Moderation: Well-Being Inequality during Episodes of Income Growth'. *Happiness and Economic Growth: Lessons from Developing Countries*, 32–139.

———. 2016. 'Economic Growth Evens Out Happiness: Evidence from Six Surveys'. *Review of Income & Wealth* 62 (3): 405–19. https://doi.org/10.1111/roiw.12190.

DIW. 2016. 'GSOEP 2015 – Erhebungsinstrumente 2015 (Welle 32) Des Sozio-Oekonomischen Panels: Personenfragebogen'. *SOEP Survey Papers Series A - Survey Instruments (Erhebungsinstrumente)*.

Ferrer-i-Carbonell, Ada, and Paul Frijters. 2004. 'How Important Is Methodology for the Estimates of the Determinants of Happiness?' *The Economic Journal* 114 (497): 641–659.

Gould, William, Jeffrey S. Pitblado, and Brian Poi. 2010. *Maximum Likelihood Estimation with Stata*. 4th ed. College Station, Tex: Stata Press.

Gupta, Prashant, Tapas Mishra, Nigel O'Leary, and Mamata Parhi. 2015. 'The Distributional Effects of Adaption and Anticipation to Ill Health on Subjective Wellbeing'. *Economics Letters* 136 (November): 99–102. https://doi.org/10.1016/j.econlet.2015.09.010.

Hurst, Erik, Geng Li, and Benjamin Pugsley. 2013. 'Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed'. *The Review of Economics and Statistics* 96 (1): 19–33. https://doi.org/10.1162/REST_a_00363.

Kaiser, Caspar, and Maarten CM Vendrik. 2019. 'Different Versions of the Easterlin Paradox: New Evidence for European Countries'. In *The Economics of Happiness*, edited by Mariano Rojas, 27–55. Springer.

Kapteyn, Arie. 1977. 'A Theory of Preference Formation'. Tilburg University. https://research.tilburguniversity.edu/en/publications/a-theory-of-preference-formation-2.

King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. 'Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research'. *American Political Science Review* 98 (1): 191–207. https://doi.org/10.1017/S000305540400108X.

Knabe, Andreas, Steffen Rätzel, Ronnie Schöb, and Joachim Weimann. 2010. 'Dissatisfied with Life but Having a Good Day: Time-Use and Well-Being of the Unemployed'. *The Economic Journal* 120 (547): 867–889.

Levinson, Arik. 2012. 'Valuing Public Goods Using Happiness Data: The Case of Air Quality'. *Journal of Public Economics* 96 (9): 869–80. https://doi.org/10.1016/j.jpubeco.2012.06.007.

Luechinger, Simon. 2009. 'Valuing Air Quality Using the Life Satisfaction Approach'. *The Economic Journal* 119 (536): 482–515.

Oswald, Andrew J. 2008. 'On the Curvature of the Reporting Function from Objective Reality to Subjective Feelings'. *Economics Letters* 100 (3): 369–72. https://doi.org/10.1016/j.econlet.2008.02.032.

Parducci, Allen. 1995. *Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications*. Happiness, Pleasure, and Judgment: The Contextual Theory and Its Applications. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Rojas, Mariano. 2007. 'A Subjective Well-Being Equivalence Scale for Mexico: Estimation and Poverty and Income-Distribution Implications'. *Oxford Development Studies* 35 (3): 273–293.

Schneider, Bruce, Scott Parker, Dan Ostrosky, David Stein, and Gary Kanow. 1974. 'A Scale for the Psychological Magnitude of Number'. *Perception & Psychophysics* 16 (1): 43–46. https://doi.org/10.3758/BF03203247.

Schröder, Carsten, and Shlomo Yitzhaki. 2017. 'Revisiting the Evidence for Cardinal Treatment of Ordinal Variables'. *European Economic Review* 92 (February): 337–58. https://doi.org/10.1016/j.euroecorev.2016.12.011.

Shepard, Roger N. 1981. 'Psychological Relations and Psychophysical Scales: On the Status of "Direct" Psychophysical Measurement'. *Journal of Mathematical Psychology* 24 (1): 21–57. https://doi.org/10.1016/0022-2496(81)90034-1.

Stevenson, Betsey, and Justin Wolfers. 2008. 'Happiness Inequality in the United States'. *The Journal of Legal Studies* 37 (S2): S33–S79.

Studer, Raphael. 2012. 'Does It Matter How Happiness Is Measured? Evidence from a Randomized Controlled Experiment'. *Journal of Economic and Social Measurement* 37 (4): 317–36. https://doi.org/10.3233/JEM-120364.

Van Praag, Bernard M. S. 1971. 'The Welfare Function of Income in Belgium: An Empirical Investigation'. *European Economic Review* 2 (3): 337–69.

———. 1991. 'Ordinal and Cardinal Utility: An Integration of the Two Dimensions of the Welfare Concept'. *Journal of Econometrics* 50 (1–2): 69–89.

Van Praag, Bernard M. S., and Ada Ferrer-i-Carbonell. 2008. *Happiness Quantified: A Satisfaction Calculus Approach*. 2nd ed. Oxford University Press.

Vendrik, Maarten C. M., and Geert B. Woltjer. 2007. 'Happiness and Loss Aversion: Is Utility Concave or Convex in Relative Income?' *Journal of Public Economics* 91 (7–8): 1423–1448. https://doi.org/10.1016/j.jpubeco.2007.02.008.

Williams, Richard. 2010. 'Fitting Heterogeneous Choice Models with Oglm'. *The Stata Journal* 10 (4): 540–567.

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics*. 4th ed. South Western Cengage Learning.

Appendix A Further details on applying equations (8)-(11) to the Easterlin Paradox example

This brief appendix extends and gives further details on the empirical analysis of Section 2.4.

For the rank-order scale $(1, 2, 3)$ of hr_i, Equation (11) implies $\hat{\beta} = -\hat{\beta}_{d1} - \hat{\beta}_{d2}$. This corresponds to Equation (1) because $\partial s_3/\partial X = -\partial(s_1 + s_2)/\partial X$ in that equation. Hence, regression model (9) for hr_i when applied to data on the Easterlin Paradox example of the main text yields a negative estimate $\hat{\beta} = -0.028$ of the effect of $\ln GDPpc$. This is equal to $-\hat{\beta}_{d1} - \hat{\beta}_{d2} = 0.025 - 0.054$ up to a rounding error. This estimate is equal to the coefficient in the regression of mean happiness $E(ht)$ in Table 1, but has a slightly smaller standard error.

As was the case in Section 2.1, there does not exist any hr scale which is concave enough to yield a statistically significant positive coefficient of $\ln GDPpc$ at the 5% or 10% level. At the infinitely strongly concave hr scale $(1,2,2)$ we get a positive coefficient of 0.025 of $\ln GDPpc$ with a slightly higher p value of 0.14 than for $\widetilde{\widetilde{\mathrm{E}(ht)}}$ in Table 1. This coefficient is exactly the opposite of the coefficient in the regression of $hd_{1,i}$ (with the same p value). Thus, the Easterlin Paradox for the USA can again not be rejected with any hr scale.

Finally, the hr scale which is just convex enough to yield a 5% significant negative coefficient of $\ln GDPpc$ is now given by $(1, 2, 4.72)$ (found numerically). This is less extreme than the scale for $\widetilde{\widetilde{\widetilde{\mathrm{E}(ht)}}}$ in Table 1 because of smaller standard errors. These clustered standard errors are likely downwardly biased because of a too low number (26) of clusters. Analogously, heteroscedasticity-robust standard errors in Table 1 tended to be smaller than the reported ordinary standard errors because of a too low number of observations (26), and hence were not used.

Appendix B Further details on implications of allowing for variation within response categories

At the start of Section 3.1, we note that Equation (2) can be extended so as to allow for changes in expectations of ht within response categories. Such a more flexible equation is given by:

$$\frac{\partial \widetilde{E(ht)}}{\partial X} = (\tilde{h}_3 - \tilde{h}_2)\frac{\partial s_3}{\partial X} + \frac{\partial(\tilde{h}_3 - \tilde{h}_2)}{\partial X}s_3 - (\tilde{h}_2 - \tilde{h}_1)\frac{\partial s_1}{\partial X} - \frac{\partial(\tilde{h}_2 - \tilde{h}_1)}{\partial X}s_1. \qquad \text{(A1)}$$

In the main text we argue that the value of $E(ht|hr = 1)$ likely rose with increases in $lnGDPpc$. Contrariwise $E(ht|hr = 3)$ likely declined with increases in $\ln GDPpc$. The unobserved derivatives $\partial(\tilde{h}_3 - \tilde{h}_2)/\partial X$ and $\partial(\tilde{h}_2 - \tilde{h}_1)/\partial X$ are therefore likely to be both negative.[60] This would make the sum of the first two terms on the right-hand side of Equation (A1) more negative than the first term on the right-hand side of Equation (2). Likewise, the sum of the last two terms in Equation (A1) will be more positive than the last term in Equation (2).

An implication of equation (A1) is that hr scales for which the effect of X on mean happiness is zero cannot be obtained from Equation (A1) by simply factoring out the ratio $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2)$ as in Equation (2). However, without loss of generality, we can simplify

[60] The increase in $\ln GDPpc$ may also have led to a change in $E(ht|hr = 2) = h_2$. However, this change is likely to be less strong than the rise in h_1 and the decline in h_3 because of diminishing marginal happiness of $\ln GDPpc$.

Equation (A1) by setting $\tilde{h}_2 - \tilde{h}_1$ equal to one (as in Table 1) and solving for $\tilde{h}_3 - \tilde{h}_2$. Setting $\partial\widetilde{E(ht)}/\partial X = 0$, this yields

$$\tilde{h}_3 - \tilde{h}_2 = \frac{\frac{\partial s_1}{\partial X} + \frac{\partial(\tilde{h}_2 - \tilde{h}_1)}{\partial X} s_1 - \frac{\partial(\tilde{h}_3 - \tilde{h}_2)}{\partial X} s_3}{\frac{\partial s_3}{\partial X}}. \qquad (A2)$$

Note that $\partial(\tilde{h}_2 - \tilde{h}_1)/\partial X * s_1$ in this expression is likely to be negative, while $-\partial(\tilde{h}_3 - \tilde{h}_2)/\partial X * s_3$ is likely to be positive. It is therefore ambiguous whether $\tilde{h}_3 - \tilde{h}_2$ must be smaller or larger than the ratio $(\partial s_1/\partial X)/(\partial s_3/\partial X)$ in order to obtain $\partial\widetilde{E(ht)}/\partial X = 0$. As stated in the main text, it is hence also ambiguous whether, compared to the estimates of Table 1, allowing for variation within response categories increases or decreases the required ratio $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2)$ at which the effect of $\ln GDPpc$ on mean happiness becomes zero.

As a second implication of Equation (A1), we can assess the likelihood that the effect of $\ln GDPpc$ on mean happiness becomes significantly positive in the limit for $\tilde{h}_3 - \tilde{h}_2 \to 0$, so for $(\tilde{h}_2 - \tilde{h}_1)/(\tilde{h}_3 - \tilde{h}_2) \to \infty$ (see the estimate for $\widetilde{\widetilde{E(ht)}}$ in Table 1). According to Equation (A1) and assuming that $\lim_{\tilde{h}_3 - \tilde{h}_2 \to 0} \partial(\tilde{h}_3 - \tilde{h}_2)/\partial \ln GDPpc * s_3$ is negligibly small, $\partial\widetilde{E(ht)}/\partial \ln GDPpc$ converges to $-\partial s_1/\partial \ln GDPpc - \partial(\tilde{h}_2 - \tilde{h}_1)/\partial \ln GDPpc * s_1$. Given the arguments above and in the main text, $-\partial(\tilde{h}_2 - \tilde{h}_1)/\partial \ln GDPpc * s_1$ in this expression is likely to be positive and likely to be positively correlated with the positive term $-\partial s_1/\partial \ln GDPpc$. Therefore, allowing for variation within response categories makes it more likely that the effect of $\ln GDPpc$ on mean happiness becomes significantly positive in the limit for $\tilde{h}_3 - \tilde{h}_2 \to 0$.

Appendix C Alternative derivation of reversal condition (16)

The factor $\pm c e^{c(\alpha_l + \beta_l X_i + \boldsymbol{\gamma}_l' \mathbf{z}_i)}$ in front of the integral in Equation (19) does not switch sign for any c. We therefore just solve for c in

$$\int_{-\infty}^{\infty} e^{c\sigma_i \epsilon_i} (\beta_l + \beta_s \sigma_i \epsilon_i) \varphi(\epsilon_i) d\epsilon_i = 0.$$

Expanding the integral in this equation yields

$$\beta_l \int_{-\infty}^{\infty} e^{c\sigma_i \epsilon_i} \varphi(\epsilon_i) d\epsilon_i + \beta_s \sigma_i \int_{-\infty}^{\infty} e^{c\sigma_i \epsilon_i} \epsilon_i \varphi(\epsilon_i) d\epsilon_i = 0.$$

The first integral equals $E(e^{c\sigma_i \epsilon_i}) = e^{0.5 c^2 \sigma_i^2}$. The second integral ($I$) can be evaluated using integration by parts. Note that $\epsilon_i \varphi(\epsilon_i) = \epsilon_i (2\pi)^{-0.5} e^{-0.5 e^2} = -\varphi'(\epsilon_i)$, and let $u = e^{c\sigma_i \epsilon_i}$ and $v'(\epsilon_i) = \epsilon_i \varphi(\epsilon_i)$. Hence, $u'(\epsilon_i) = e^{c\sigma_i \epsilon_i} c\sigma_i$ and $v(\epsilon_i) = -\varphi(\epsilon_i)$, yielding

$$I = \int_{-\infty}^{\infty} e^{c\sigma_i \epsilon_i} \epsilon_i \varphi(\epsilon_i) d\epsilon_i = -e^{c\sigma_i \epsilon_i} \varphi(\epsilon_i)|_{-\infty}^{\infty} + c\sigma_i \int_{-\infty}^{\infty} e^{c\sigma_i \epsilon_i} \varphi(\epsilon_i)\, d\epsilon_i.$$

Evaluating the first term at either limit of integration leads to

$$\lim_{\epsilon_i \to \pm\infty} -e^{c\sigma_i \epsilon_i} \varphi(\epsilon_i) = \lim_{\epsilon_i \to \pm\infty} -e^{c\sigma_i \epsilon_i} (2\pi)^{-0.5} e^{-0.5\epsilon_i^2} = -(2\pi)^{-0.5} \lim_{\epsilon_i \to \pm\infty} e^{c\sigma_i \epsilon_i - 0.5\epsilon_i^2} = 0.$$

Hence, $I = c\sigma_i E(e^{c\sigma_i \epsilon_i}) = c\sigma_i e^{0.5 c^2 \sigma_i^2}$. We therefore obtain

$$\beta_l e^{0.5c^2\sigma_i^2} + \beta_s c\sigma_i^2 e^{0.5c^2\sigma_i^2} = (\beta_l + \beta_s c\sigma_i^2)e^{0.5c^2\sigma_i^2} = 0$$

Solving for c yields

$$c = -\frac{\beta_l}{\sigma_i^2\beta_s} = -\frac{\beta_l}{e^{2(\alpha_s+\beta_s X_i+\gamma_s{}'z_i)}\beta_s},$$

which is condition (16).

Appendix D An argument from psychophysics

Psychophysicists analyse how subjective intensities across different "modalities" relate to intensities of objective stimuli. To do so, respondents are often asked to match a given stimulus s^m of modality m (e.g. a sound with a given physical intensity), which is associated with a subjective intensity ψ^m (i.e. that sound's subjective loudness), with another stimulus of a different modality s^n (e.g. the objective luminosity of a lamp) with intensity ψ^n (i.e. that lamp's subjective brightness). Such procedures are known as "cross-modality matching" (Shepard 1981).

Answers to happiness questions can be placed in this framework. Note that a person's true happiness ht represents a subjective intensity ψ^{ht}.[61] Furthermore, the R numbered response options in any happiness question can be viewed as physical number stimuli $s_1^{num},\dots,s_R^{num}$, associated with subjective intensities $\psi_1^{num},\dots,\psi_R^{num}$. Although surveys are not explicitly framed as cross-modality matching tasks, respondents may attempt to match the subjective intensity of each numbered response option with their felt happiness intensity. When choosing a response option with number k, a respondent at time t may thus minimize the difference $\psi_k^{num} - \psi_t^{ht}$. In turn, this implies that when a respondent initially answered e.g. "6" and now answered e.g. "8", we can infer that the difference in happiness intensities is approximately equal to the difference in the subjective intensities of the number stimuli "6" and "8", i.e. that $\psi_{t=2}^{ht} - \psi_{t=1}^{ht} \approx \psi_{k=8}^{num} - \psi_{k=6}^{num}$. Therefore, if we were to identify the differences between subjective intensities of numbers, we can infer the differences in psychological magnitudes of ht.

Banks & Coleman (1981) tried to experimentally infer the function that relates objective with subjective magnitudes of a given finite set of numbers.[62] It is not feasible to directly ask how certain numbers "feel", since when e.g. asked whether the difference between the numbers "6" and "8" is larger than difference between the numbers "3" and "4", respondents would likely apply formal arithmetic rules. To circumvent this problem, Banks & Coleman utilized the following procedure: Respondents were given multiple series of ten different integers I on a known interval (e.g. from 1 to 1000). Each series was generated by a function of the form $I = ax^b$, with $x = 1, 2, \dots, 10$; $b = 0.1, 0.3, 0.5, 0.7, 0.9, 1.1$ (note that $b = 1$ would yield a linear function) and with a chosen such that the minimum and maximum of the series are as close as possible to the interval boundaries. Within each series, b was held constant while using every value of x. The resulting integers were

[61] Happiness is unusual in that there is no single physical stimulus s^{ht} corresponding to a particular level of ψ^{ht}.

[62] Which is a function satisfying $f(s_k^{num}) = \psi_k^{num}$. We only report Banks & Coleman's results for cases with known bounded intervals within which numbers can occur. This is because only these results are directly relevant to typical response scales in happiness research. Banks & Coleman also present results for intervals without a known upper bound. In such cases, they find that power functions with exponents between 0.3 and 0.5 (i.e. concave functions) are a much better approximation of the subjective intensities of numbers. These results are in line with those of Schneider et al. (1974) as well as Banks & Hill (1974) who use similar methods.

then randomly shuffled. Respondents were next asked to rate how "random" these series of numbers felt. Under the assumption that respondents took "random" to mean "sampled from a uniform distribution", we should expect respondents to give series where the distribution of subjective intensities more closely resembled a uniform distribution to receive a higher subjective "randomness" rating.[63] It turns out that functions with exponents $b = 0.9$ or $b = 1.1$ received the highest mean randomness rankings across respondents. This suggests that the subjective intensity of numbers within a known bounded range is close to linear in the objective magnitude of such numbers. In a separate experimental setup, Banks & Coleman also asked respondents to rapidly think of 25 random numbers on some bounded interval (either 1-10, 10-99, or 100-999). Respondents' answers were put in rank order and means for each rank were calculated across respondents. OLS regressions of these means on their rank yielded R-squared statistics between 0.974 (for the interval 10-99) and 0.995 (for the interval 1-10). These results also suggest a linear relation between objective values of numbers and their subjective intensities.

Given the aforementioned arguments, if the subjective intensity of numbers increases linearly with their objective magnitude on a bounded interval, we have further reason to believe that rank-order hr and ht also relate linearly.

[63] This assumes that how random a series of numbers feels, only depends on the subjective intensities of the numbers in the series, and not on their objective magnitudes. Since felt randomness is a subjective perception itself, we find this assumption reasonable.

Appendix E Additional tables

Table A1. Cumulative response shares for happiness and life satisfaction in ESS and WVS for European countries

WVS (Happiness)		ESS (Happiness)			ESS (Life Satisfaction)		WVS (Life Satisfaction)	
hr	% share (cumulative)	hr	% share (cumulative)	$\overline{hr}$ after collapse	hr	% share (cumulative)	hr	% share (cumulative)
1	2.45 (2.45)	0	0.97 (0.97)	0.52	0	3.26 (3.26)	1	2.40 (2.40)
		1	1.04 (2.01)		1	2.15 (5.41)	2	1.99 (4.39)
2	13.06 (15.51)	2	2.11 (4.12)	3.23	2	3.37 (8.78)	3	4.16 (8.55)
		3	3.88 (8.00)		3	6.10 (14.88)	4	4.55 (13.11)
		4	4.47 (12.47)		4	5.92 (20.80)	5	11.93 (25.04)
3	58.98 (74.50)	5	14.60 (27.06)	6.79	5	14.78 (35.58)	6	10.83 (35.87)
		6	9.24 (36.30)		6	9.45 (45.03)		
		7	18.70 (55.00)		7	16.25 (61.29)	7	18.77 (54.64)
		8	24.14 (79.13)		8	21.24 (82.52)	8	25.27 (79.90)
4	25.50 (100.0)	9	11.86 (90.99)	9.43	9	9.38 (91.90)	9	11.79 (91.69)
		10	9.01 (100.0)		10	8.10 (100.0)	10	8.31 (100.0)

Note: Data from WVS wave 5 and ESS wave 3 (both 2006). Design and population weights applied. Countries included: France, Finland, Germany, Great Britain, The Netherlands, Norway, Poland, Romania, Russia, Slovenia, Spain, Sweden, Switzerland, Ukraine. WVS response options for happiness are labelled "Not at all happy" (=1), "Not very happy" (=2), "Rather happy" (=3), "Very happy" (=4). Extreme response options for happiness in ESS are labelled "Extremely unhappy" (=0) and "Extremely happy" (=10). Extreme response options for life satisfaction in ESS are labelled "Extremely dissatisfied" (=0) and "Extremely satisfied" (=10). Extreme response options for life satisfaction in WVS are labelled "Completely dissatisfied" (=0) and "Completely satisfied" (=10).

Table A2. Replication of Table 1 with added linear time trend

	s_1	s_2	s_3	$\mathrm{E}(ht)$
ln*GDPpc*	-0.316**	0.016	0.299*	0.614**
	(0.130)	(0.181)	(0.162)	(0.232)
Year	0.006**	0.001	-0.007**	-0.013*
	(0.003)	(0.004)	(0.003)	(0.005)

Note: * p < 0.10; ** p < 0.05, *** p < 0.01. Rows for ln*GDPpc* and year denote regression coefficients with ordinary standard errors in parentheses. $\mathrm{E}(ht)$ holds for a rank-order coding of hr.

Table A3. Relabelling condition for the pooled OLS model

	(1) hr ≤ 0	(2) hr ≤ 1	(3) hr ≤ 2	(4) hr<=3	(5) hr ≤ 4	(6) hr ≤ 5	(7) hr ≤ 6	(8) hr ≤ 7	(9) hr ≤ 8	(10) hr ≤ 9
					No Controls					
Log household income	-0.006***	-0.010***	-0.021***	-0.039***	-0.061***	-0.145***	-0.180***	-0.166***	-0.066***	**0.003****
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	**(0.001**)
Unemployed	0.015***	0.025***	0.052***	0.100***	0.150***	0.258***	0.295***	0.250***	0.102***	0.028***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.001)
Married	-0.002***	-0.004***	-0.009***	-0.016***	-0.023***	-0.035***	-0.040***	-0.043***	-0.013***	-0.004***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Children	-0.001**	-0.002***	-0.004***	-0.009***	-0.014***	-0.032***	-0.042***	-0.039***	-0.029***	-0.003*
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Disability	0.010***	0.019***	0.037***	0.066***	0.096***	0.176***	0.196***	0.166***	0.077***	0.014***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
					Full Controls					
Log household income	-0.003***	-0.006***	-0.014***	-0.027***	-0.045***	-0.110***	-0.143***	-0.143***	-0.068***	-0.009***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.001)
Unemployed	0.013***	0.022***	0.044***	0.083***	0.121***	0.190***	0.208***	0.160***	0.058***	0.017***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.002)	(0.001)
Married	-0.003***	-0.005***	-0.011***	-0.021***	-0.030***	-0.053***	-0.063***	-0.065***	-0.031***	-0.008***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Children	**0.000**	-0.001	-0.002**	-0.005***	-0.010***	-0.020***	-0.029***	-0.035***	-0.023***	-0.008***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.001)
Disability	0.009***	0.017***	0.034***	0.060***	0.088***	0.149***	0.164***	0.149***	0.071***	0.026***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
Observations	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Cells in bold have opposite sign, implying possibility of reversal.

Table A4. Relabelling condition for the FE model

	(1) hr ≤ 0	(2) hr ≤ 1	(3) hr ≤ 2	(4) hr<=3	(5) hr ≤ 4	(6) hr ≤ 5	(7) hr ≤ 6	(8) hr ≤ 7	(9) hr ≤ 8	(10) hr ≤ 9
					No Controls					
Log household income	-0.003***	-0.006***	-0.011***	-0.020***	-0.030***	-0.052***	-0.059***	-0.045***	-0.008***	**0.005*** **
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	**(0.001)**
Unemployed	0.010***	0.016***	0.035***	0.065***	0.093***	0.131***	0.143***	0.104***	0.033***	0.013***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	(0.001)
Married	-0.002***	-0.004***	-0.010***	-0.017***	-0.023***	-0.036***	-0.040***	-0.033***	-0.009***	**0.002**
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.003)	**(0.002)**
Children	-0.000	-0.001**	-0.003***	-0.007***	-0.011***	-0.012***	-0.013***	-0.012***	-0.003+	-0.005***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Disability	0.005***	0.011***	0.021***	0.039***	0.061***	0.095***	0.104***	0.092***	0.045***	0.020***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)	(0.002)
					Full Controls					
Log household income	-0.002***	-0.005***	-0.010***	-0.019***	-0.030***	-0.054***	-0.067***	-0.065***	-0.033***	-0.009***
	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.001)
Unemployed	0.010***	0.016***	0.034***	0.064***	0.091***	0.129***	0.140***	0.103***	0.035***	0.015***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	(0.001)
Married	-0.001*	-0.002***	-0.007***	-0.013***	-0.018***	-0.028***	-0.035***	-0.038***	-0.023***	-0.004**
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)
Children	-0.000	-0.000	-0.001	-0.004**	-0.005***	-0.005**	-0.004+	**0.000**	**0.008*** **	**0.003** **
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	**(0.003)**	**(0.003)**	**(0.002)**
Disability	0.005***	0.009***	0.016***	0.028***	0.044***	0.070***	0.072***	0.053***	0.011***	**-0.001**
	(0.001)	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.003)	**(0.002)**
Observations	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999	557,999

Note: * p<0.10, ** p<0.05, *** p<0.01. Clustered (by respondent) standard errors in parentheses. Cells in bold have opposite sign, implying possibility of reversal.

Appendix F Additional figures

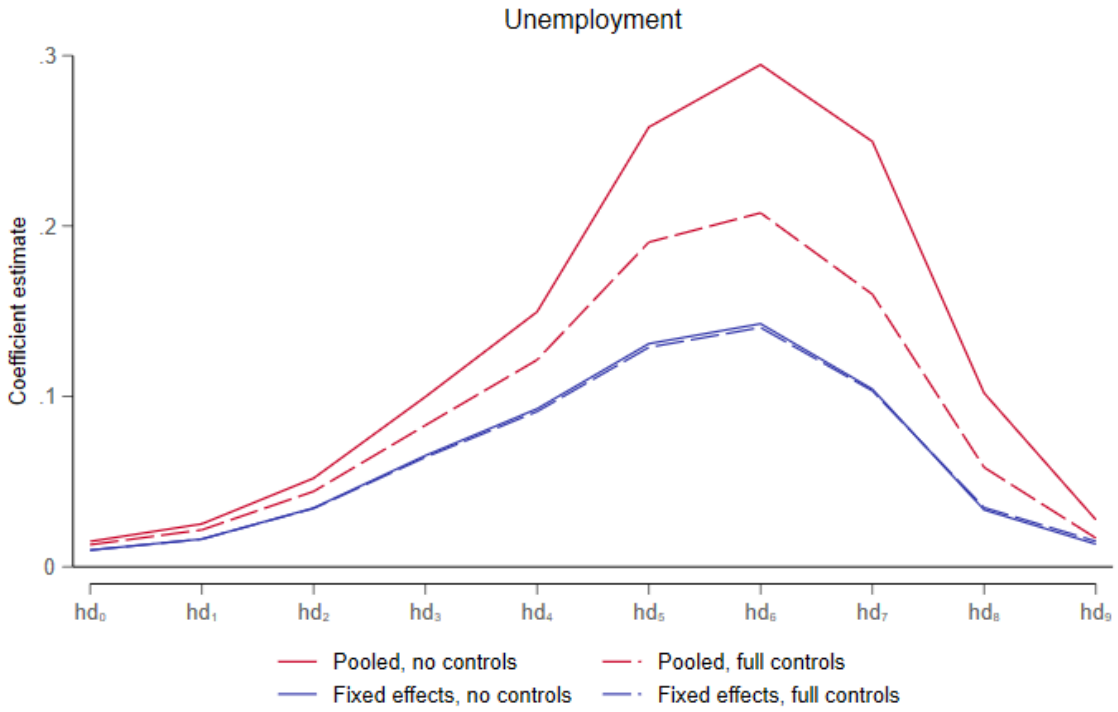

Figure A1. Coefficients estimates of unemployment for each regression of $hd_{k,it}$, corresponding to Appendix Tables A3 and A4.

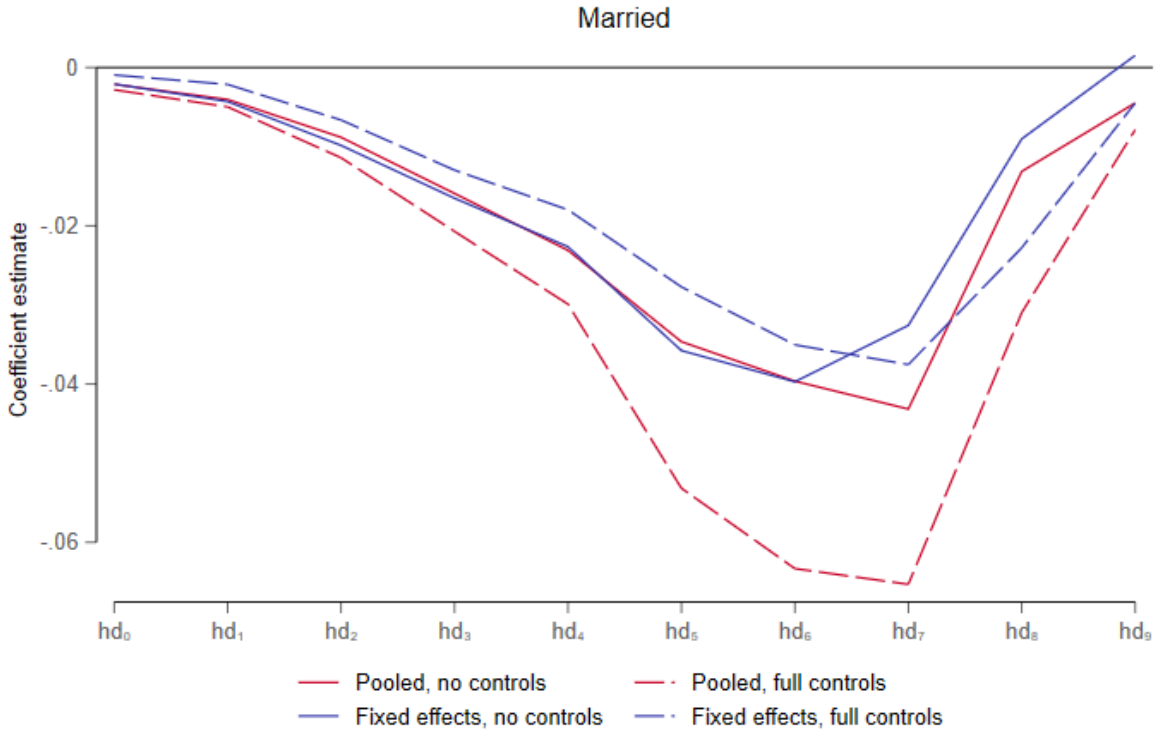

Figure A2. Coefficients estimates of marriage for each regression of $hd_{k,it}$, corresponding to Appendix Tables A3 and A4.

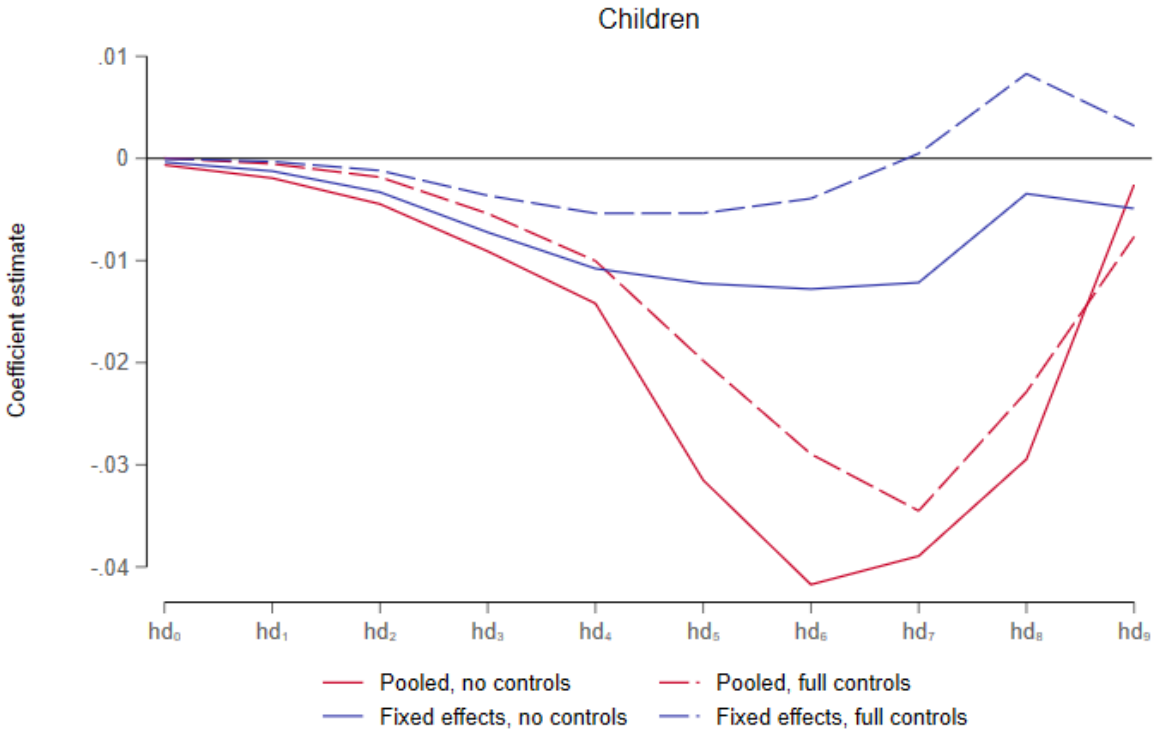

Figure A3. Coefficients estimates of children for each regression of $hd_{k,it}$, corresponding to Appendix Tables A3 and A4.

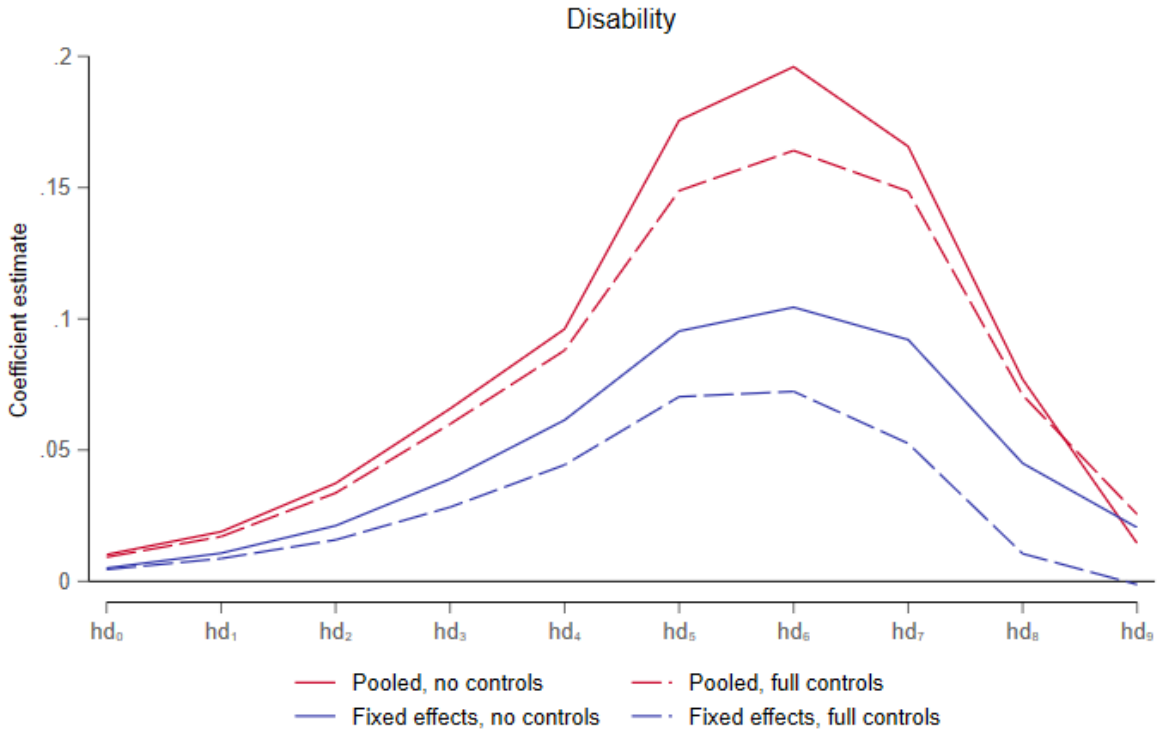

Figure A4. Coefficients estimates of disability for each regression of $hd_{k,it}$, corresponding to Appendix Tables A3 and A4.